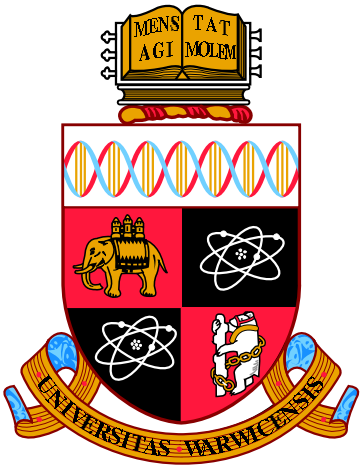


Resource Allocation for Multi-Sensory Virtual Environments

by

Efstratios Doukakis

Thesis

Submitted to the University of Warwick for the degree of

Doctor of Philosophy

Warwick Manufacturing Group

September 2016

THE UNIVERSITY OF WARWICK

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to all those who gave me the possibility to complete this thesis. First of all, my supervisors Kurt and Alan. Both of them gave me guidance and courage to keep going while they supported my application for extra funding towards the end of my PhD. Saying "thank you" is an understatement.

I would like also to thank the research fellows of the visualisation group. Carlo and Tom were always eager to provide assistance to every PhD student without even being asked to do so. I am also very lucky for meeting and spending time with Rosella, Debmalya, Ali, Pinar, Jon, Amar, Josh, Tim and Emmanuel. Thanks for everything. Ratnajit, thanks for the stimulating discussions and the fun we have had the last four years. I was very fortunate to come across many people during this PhD. Doctor Andrew Garmory and Professor Martin Passmore helped me a lot with my odour transport simulations during my visits to Loughborough university. Ahmed al Makky, who inspired me to get started with OpenFoam and implement turbulence modelling in my simulations. I would like also to thank Doctor Derrick Watson for our discussions about psychophysics procedures. Many of the ideas presented in chapter 9 have come about due to discussions with him. Many thanks also to my examiners, Professor Caroline Meyer of Warwick University and Doctor Kirsten Cater of Bristol University for making the viva very pleasant.

I would like to thank my parents and sister for their motivation and encouragement. Joanna, I would not have made this without your love and support. This thesis is dedicated to you.

Declarations

Peer-reviewed journal papers:

- Doukakis E., Debattista K., Bashford-Rogers T., Harvey C., Chalmers A.: Audio-Visual Resource Allocation for Bi-Modal Virtual Environments, Computer Graphics Forum. (conditionally accepted) 2016.

Name: Efstratios Doukakis **Date:** 30 September 2016

Abstract

Fidelity is of key importance if virtual environments are to be used as authentic representations of real life scenarios. However, simulating the multitude of senses that comprise the human sensory system is a computationally challenging task. With limited computational resources it is essential to distribute these carefully in order to simulate the most ideal perceptual experience for the user.

This thesis investigates this balance of resources across multiple scenarios where combined audio, visual and olfactory cues are delivered to the user. Starting with bi-modal virtual environments where audio and visual stimuli are delivered to the users, a subjective experimental study, denoted as ***E1***, was undertaken where participants ($N = 35$) allocated five fixed resource budgets for adjusting the quality of the displayed graphics and acoustics stimuli. In this experiment, increasing the quality of one of the sensory stimuli decreased the quality of the other. Findings demonstrate that participants allocate more resources to graphics, however as the computational budget is increased the allocation ratio between graphics and acoustics decreases significantly. Based on the results, an audio-visual quality prediction model is proposed and successfully validated against previously untested budgets and scenarios.

The introduction of realistic olfactory stimuli is considered necessary if multisensory virtual environments are to be used as genuine representations of real life experiences. The estimation and delivery of smell impulses includes many challenges and significantly differs from the methods used for computing and displaying auditory and visual cues. Furthermore, the absence of a quality metric for assessing olfactory stimuli makes the introduction of an olfactory quality scale in the resource allocation framework significantly challenging.

The work presented in this thesis investigates whether better spatial discretisation of the computational domain, a frequently used technique for increasing numerical stability in fluid transport simulations, can be used as a successful smell quality metric that can elicit a perceptual impact to the users of the virtual environment. In this context, better spatial discretisation levels are evaluated based on the estimation of the Just Noticeable Difference (JND) threshold for smell intensity using an experimental study ($N = 20$) and implemented in two phases. This experiment is denoted as ***E2*** throughout this thesis. Findings demonstrate that the JND threshold is larger than the concentration differences given by progressively accu-

rate smell transport simulations. This outcome enables computational savings from avoiding exhaustive smell transport simulations that provide no perceptual benefit to the user.

Having considered the limitations associated with assessing smell impulses in terms of quality, a third experimental study is proposed, denoted as ***E3***, and is exploited for resource allocation in tri-modal virtual environments. The experimental layout of ***E3*** ($N = 25$) builds on the framework proposed in ***E1*** including the delivery of physically accurate smell impulses to the user. The display of the smell bursts is implemented in a binary fashion (two levels or ON/OFF smell) along with the quality levels for the senses of vision and hearing as selected in ***E1***. The smell concentration level presented in this experiment follows from the results of the JND threshold estimation for odour concentration presented in the psychophysics study ***E2***.

In conclusion, the research presented, shows that human preference criteria can be fully exploited in the design and delivery of multi-sensory virtual experiences. Experimental data can be used to derive computationally inexpensive prediction models that direct resource allocation in rendering pipelines where diverse sensory stimuli are simulated and delivered to the users.

Keywords: Multi-sensory virtual experiences, virtual reality, Bi-modal, Tri-modal, Olfaction.

Chapter 1

Introduction

Virtual Environments (VEs) provide the opportunity to simulate a wide range of applications, from training to entertainment, in a safe and controlled manner. VEs intend to provide the experience of being in a real place where interaction with objects or characters is possible. A primary challenge of VEs is to provide realistic representations of real world environments. For applications which require high levels of authenticity, the VEs need to provide multiple, physically accurate sensory stimuli. The existence of multiple senses in the VE context expands the number of potential applications while it increases the level of immersion [HB96, LVK02]. Examples include, but are not limited to, commercial vehicle simulators [YSYT94, FMT11], video games [RLC+07, GBW+09] and applications in the prognosis and treatment of various human disorders in healthcare [Riv03, TS10] and neuropsychology [RBvdZ+00, Mor04]. Examples of multi-sensory VEs are given in figure 1.1.

Figure 1.1: Examples of multi-sensory virtual environments. **Left:** Driving simulator [PfSI12], **Middle:**Virtual infirmary [Vir16], **Right:** Video games [Ger99].

Research studies and practical applications have mainly focused on including stimuli coming from the senses of sight and hearing in the VEs as these are considered more dominant compared to the other senses of the Human Sensory System (HSS) [Sto98]. Audio and visual stimulations are very important for obtaining a full cognition of the virtual world as they allow localisation of its various features

(e.g. boundaries, objects, etc.) and provide an understanding of its shape, scale and surrounding properties (e.g. surface materials).

Other senses have recently been introduced in bi-modal frameworks as a mean of increasing the level of fidelity and sense of presence in the VEs. Examples in the literature include haptics, feel and smell cues [MC97, HS13]. The significance of the sense of olfaction, in particular, might not be initially obvious, but it affects how humans feel, think and behave in real life. Furthermore, there is evidence that people tend to recall memories associated with a perceived odour [HE96, SP04, PSNG06]. Olfactory cues have only recently been exploited in multi-sensory rendering frameworks either with the objective of reducing the computational complexity of displayed visual stimuli [RCHR07, BCB+09] or for enhancing saliency estimation models under the effect of congruent smell stimuli in the virtual scene [HBRDC11].

The computation and delivery of multi-modal VEs in high fidelity requires significant computing resources. Previous work has shown that humans are not able to attend to all the sensory stimuli at the same time in real life situations [Sto98, TN03a]. Such cross-modal effects have mostly been exploited in bi-modal VEs where, usually combined audio and visual stimuli are delivered to the user. Perceptual interactions between the senses have been used to compute and deliver lower quality visual [CCL02, RFWB07, HDAC10] or auditory [TGD04, MBT+07] stimuli without the quality compromise being consciously perceived by the average user. Furthermore, many examples in the literature have shown that the perceived quality of a sensory stimulus can be significantly increased under the synchronous effect of a different sensory cue. Interactions of this kind have mainly been used to increase the temporal or spatial quality of visuals under the effect of auditory stimulations [MDCT05a, MDCT05b, HAC08, HDAC11].

While failure to attend to certain parts of a virtual scenario has been used to reduce the computational requirements, precisely how best to allocate a fixed amount of computational resources in the simulation of multiple stimuli has not been quantified in previous studies. The work presented in this thesis examines how a budget of available resources affects the perceived fidelity of a multi-sensory virtual experience firstly in bi-modal VEs where audio and visual stimuli are delivered to the users and subsequently in tri-modal VEs where physically accurate olfactory impulses are displayed along with the other senses.

1.1 The research problem and its significance

The simulation of various sensory stimuli for computing realistic virtual experiences is still not feasible at interactive rates with the existing algorithms and hardware available today. The demand of VEs in different areas of academia and industry is continuously increasing and the availability of computing resources in these areas varies. It is necessary, therefore, to exploit the available resources in the best possible way for rendering the sensory cues of the VEs. For this purpose, allocating a large amount of resources, either hardware or software, to the senses that are more important for the perceived fidelity and fewer resources to the senses that do not contribute to the overall perceptual experience is considered crucial.

Multi-sensory rendering pipelines exploit interactions between the senses in order to reduce the rendering times by delivering lower quality stimuli without the users being able to discern any degradation in the quality of the overall experience. However, there is no generic way to quantify the level of this interaction so as to adjust the distribution of the resources to the senses of the VE accordingly. Existing multi-sensory rendering frameworks are solely based in heuristics that have reported and studied in perception literature for reducing the rendering times but no work has looked into quantifying this process formally.

The research problem is firstly approached in the context of two senses, namely sight and hearing. The experimental layout is later extended to accommodate the delivery of olfactory impulses in a tri-modal VE set-up. The experimental layout can be generalised to any number of different sensory stimuli for distributing the available resources.

1.2 Overall method

Picture 1.2 depicts the overall method followed for approaching the resource allocation problem based on human subjective preferences. An experimental study is conducted where participants allocate a given budget of resources to improve the quality of the different sensory stimuli in a finite number of virtual scenarios. The assigned budget corresponds to potentially quality improvements; in turn, these quality improvements in the considered senses directly correspond to a higher computational cost that is allocated from the available budget. Participants' allocation preferences are captured in the experimental data which are used in the construction of a statistical model. This model takes as input a budget size and a scenario and estimates the percentages that should be given to every sense so as the best possible

perceptual experience will be delivered to the user.

Adjusting the quality of the sensory stimuli in the experimental layout is a process that depends on many parameters. For instance, the quality of the rendered images strongly depends on the algorithm selection (ray tracing/rasterized images), complexity of the scene, material selection and many others. All these parameters adjust the quality level and affect the computation times. This thesis proposes a methodology for decoupling the resource allocation problem from these factors by using normalised costs that are independent of the algorithm selection and its parameterisation.

1.3 Research objectives and contributions

The experiments implemented in the context of this thesis have a twofold purpose. Firstly, to investigate whether participants' subjective preferences for resource allocation follow any trends under the effect of different budget sizes and scenarios. Secondly, to utilise the experimental data in actual applications by designing prediction models capable of estimating the percentage of the total budget that should be devoted to do quality improvements in each sense. As the participants allocate resources depending on the quality of the sensory stimuli, it is important to define quality metrics for every sense of the experimental framework. This thesis considers the challenges associated with the definition of metrics in each of the visual, auditory and olfactory modalities.

The contributions of this thesis can be listed as follows:

- A literature review of the field of perceptually based rendering including the description of the psychological phenomena that characterise the co-existence of multiple sensory stimuli.
- The establishment of a solid experimental methodology for resource allocation in bi-modal VE based on user's actual preferences.
- The development and validation of a statistical model for estimating audio-visual distribution for an input budget and scenario.
- A methodology for rendering physically accurate visual, auditory and olfactory stimuli in the design and delivery of virtual experiences.
- An investigation of whether better spatial discretisation of the VE's computational domain can elicit perceptual differences in the olfactory domain and

whether these differences can be used as a representative metric for assessing olfactory impulses.

- An extension of the experimental methodology to accommodate the existence of multiple senses giving as example the inclusion of smell impulses in the resource allocation framework.

- Finally, the development of a new statistical model that can estimate audio-visual allocation and additionally predict whether smell impulses should be delivered to the users.

1.4 Thesis outline

This thesis is organised as follows:

- Chapter 2 provides an introduction to the physical and perceptual properties of light, sound and smell. The chapter also explains the basic methods to deliver visual, auditory and olfactory stimuli to the users. An overview of the basic simulation pipeline for computing the three senses is also given.

- Chapter 3 presents an overview of the human perceptual phenomena that consist of the theoretical background of many selective rendering schemas. The discussion is then directed to perceptual interactions between the senses and the way they can be used to reduce computation times in cross-modal selective rendering frameworks.

- Chapter 4 describes the research methodology of the resource allocation experimental studies presented in this thesis. The objectives of the experiments ***E1***, ***E2***,***E3*** and their importance for answering the research question are given.

- Chapter 5 explains the process of computing and assessing visual stimuli in the form of rendered images. These are necessary materials for the implementation of the experimental studies ***E1***, ***E3***.

- Chapter 6 describes the methodology followed for computing and evaluating audible cues. Auditory stimuli are also part of the software materials used for conducting the studies ***E1***, ***E3***.

- Chapter 7 examines how the budget of available resources affects the perceived fidelity of an audio-visual experience in bi-modal VEs. A main experimental study (***E1***) is conducted for capturing participants' allocation preferences.

The obtained data are used in the construction of a statistical model for estimating audio-visual resource distribution. The model performance is assessed in a validation experimental study.

- Chapter 8 describes the process followed for simulating and delivering physically accurate smell impulses to the users of a VE. This chapter serves for preparing the olfactory materials needed for conducting the experimental studies ***E2*** and ***E3***.

- Chapter 9 presents a psychophysics experimental study (***E2***) for estimating the JND threshold for odour intensity. The objective of this study is to investigate whether better spatial discretisation of the computational domain in smell transport simulations is a representative metric of olfactory quality.

- Chapter 10 extends the experimental methodology of resource allocation from bi-modal to multi-modal VEs and demonstrated in a tri-modal set-up with the inclusion of the sense of olfaction along with the other senses. A new experiment (***E3***) is conducted for capturing participants' preferences. The experimental data are used in the construction of a statistical model whose performance is tested in a validation study.

- Finally, in chapter 11, conclusions and future plans for extending this work are discussed. The contribution of this research in the field of Computer Graphics is also given.

- Appendix A provides the definition of the solid angle used in chapter 5.

- Appendix B provides the basic points of linear and time invariant filters and the process of convolution needed in chapter 6.

- Appendix C provides the vector calculus definitions used in chapter 8.

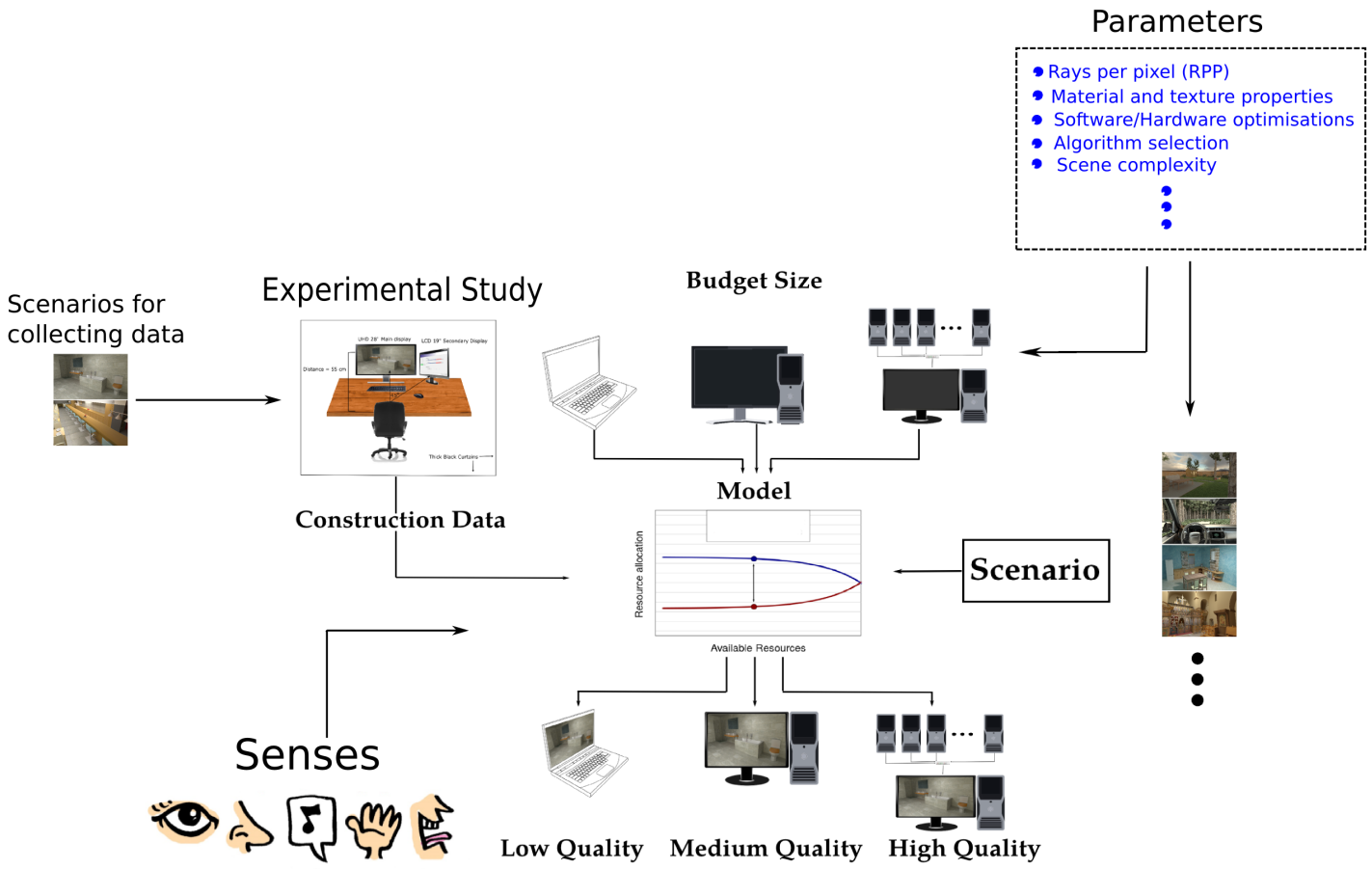


Figure 1.2: Research methodology schematic diagram.

Chapter 2

Background

This chapter introduces the fundamental properties of light, sound and smell and describes methods to deliver stimuli of these three senses. The basic terminology and concepts of this chapter are necessary for understanding the rest of this thesis.

2.1 Introduction to light

Light can be described either as a bundle of photons (corpuscular theory) or as an electromagnetic wave (wave theory). Electromagnetic waves are composed of an electric and magnetic field that fluctuate together and form a propagating wave which carry energy from one point to another. Every electromagnetic wave has a frequency f, a propagation speed u and a wavelength λ which are connected by the equation $u = f\dot{\lambda}$. The wave nature of light was used to explain macroscopic phenomena such as reflection, refraction, diffraction and interference [HRW13].

2.1.1 Physical properties of light

Speed of light

Light can travel in vacuum at a speed of $c = 3 \times 10^8$ m/s. The speed of light in air is approximately the same as in vacuum and the same value is used in practical applications. Light can travel through material media and its speed is modulated by the *index of refraction* of the media. When light propagates through media whose index of refraction is strictly higher than one its speed reduces. Air has unity index of refraction. The electromagnetic properties of the medium also affect the speed of

light. The speed of light in this case can be found using:

$$u = \frac{1}{\sqrt{\epsilon_0 \mu_0}}, \tag{2.1}$$

where μ_0 is the magnetic permeability and ϵ_0 is the electric permittivity of the material [HZZ01].

Wavelength and frequency

Electromagnetic waves can be classified depending on their wavelength. The electromagnetic spectrum encompasses the whole range of wavelengths from 10^{-11} cm (gamma rays) up to 10^3 cm (radio waves). This range contains the region of visible light where the human eye operates and perceives visual stimuli. The visible spectrum ranges from 380 nm up to 750 nm[1] or in terms of frequency between 4.3×10^{14} Hz and 7.7×10^{14} Hz. The human eye recognises light wavelengths as different colours ranging from violet (short wavelength) up to red colour (long wavelength). All the other visible colours exist in between these wavelengths.

Radiometry and photometry

Properties of light can be described either using physical quantities that involve light energy or by quantifying the perceptual impact of these physical quantities to human observers. The former way describes the area of *Radiometry* while the latter is the subject of *Photometry* in the field of Optics. Radiometric quantities include physical measurement of light properties and they are used in the description of light transport through a scene using global illumination algorithms [DBB06]. Photometric quantities express the sensitivity of the human eye at different frequencies of the visible spectrum and their calculation is based on the derivation of the corresponding radiometric quantities [SAM09]. Throughout this thesis both radiometric and photometric quantities are used for describing fundamental properties of light.

Intensity of light

The intensity of light coming from a distant source to an object can be generally approximated if the distance between the source and the object is known. Specifically, the intensity of light reaching an object is inversely proportional to the square of the distance between the object and the light source. This idealisation of the light

[1] $1\text{nm} = 10^{-9}\text{m}$.

intensity estimation does not take into account cases where light can be absorbed or scattered by the surfaces of the environment.

The radiometric quantity used for describing light intensity is known as *Irradiance* and is measured in Watts/m^2. Irradiance can be calculated as

$$E = \frac{I}{R^2}, \tag{2.2}$$

where I is the intensity of the light source (measured in Watts) and R denotes the distance between the illuminated object and the light source. Light intensity can also be described using photometric quantities (see section 2.1.3).

2.1.2 Perceptual properties of light

Just noticeable difference threshold

A useful definition when studying the perceptual impact of a physical property is that of the Just Noticeable Difference or JND threshold. It expresses the magnitude that needs to be added to a physical stimulus so as the increase to be perceptually detectable by the average human. JND thresholds can be estimated using psychophysics experiments and cannot be measured directly using scientific instruments.

Luminance and contrast

The photometric quantity that is used to measure light intensity is termed *Luminance* or *Brightness* and is measured in cd/m^2. Candela (cd) is the measurement unit used to describe how much light is emitted by a light source. Luminance expresses how much light is emitted by a light source per unit area or generally what is the amount of light that is absorbed or reflected by a region [WA01]. Contrast is the difference in luminance between an object and its surroundings and is the primary reason humans can visually perceive objects in real life or in a visual display. The JND threshold for contrast has been measured in previous research and found equal to 0.09 cd/m^2 [KP09].

At normal lighting conditions the HVS can recognise alternating black and white stripes of high contrast. The human eye can resolve this pattern by identifying 50 to 60 such black-white stripe pairs per visual angle. A typical LCD monitor is not able to represent more that 20 pairs per visual angle, on average [SAM09]. As the frequency of the alternating stripes increases the human eye starts blending the black and white colours and a uniform grey region is formed indicating that the HVS is not

anymore able to resolve the grating pattern. The ability to perceive this alternating pattern for increasing angular frequencies can be encoded using a human *Contrast Sensitivity Function* (CSF) that shows how contrast perception scales for increasing spatial frequencies in the visual field. CSFs can be estimated using psychophysical experimental studies where people's ability to detect alternating grading patterns is captured for different frequencies.

Weber's law states that the ratio between the JND threshold of any stimulus and its magnitude is constant. Specifically, Weber's law for contrast can be written in the following form:

$$\frac{\Delta L}{L_b} = k, \tag{2.3}$$

where $\Delta L = L_b - L_p$ expresses the difference in luminance between a patch (L_p) and its backround (L_b). A generalisation of Weber's law was proposed by Fechner in 1870 and expresses the perceived intensity of any stimulus as a function of its physical magnitude.This law can be written as:

$$S = \lambda \log(I), \tag{2.4}$$

where S is the perceived intensity of any stimulus and I is its physical magnitude. The constant λ depends on the stimulus under consideration [SAM09].

Colour perception

The human retina absorbs large collections of photons that have different wavelengths (or different energies). The incoming light information is translated into electrical signals that are processed by the Human Visual System (HVS) and perceived as colours of different intensities. The colours of the visible spectrum that have a single wavelength value are termed as *Spectral colours*. Mixtures of different wavelength values can result to colours that elicit the same visual perception as a monochromatic spectral colour [Hun04].

In the past, many studies have described colour matching experiments where human subjects were asked to use different combinations of primary colours in order to match given coloured patches. The experimental results serve as an approximation to the average human observer's response to colour and have been standardised. The Commission Internationale d'Eclairage (CIE) [CIE24, Ste88] has defined three monochromatic primaries at wavelengths 435.8, 546.1 and 700 nm that were used to match colours of complex spectral compositions by combining different amounts of each given primary colour. The exact values required for receiving a perceptual

match can be found by using the colour matching functions which result from the colour matching experimental results. The colour matching functions are used to derive tristimulus values, denoted as R, G and B, that encode the HVS's response to the spectral composition of the incoming light. Using the tristimulus values, any colour C can be written as:

$$C = R\mathbf{R} + G\mathbf{G} + B\mathbf{B},$$

where **R**, **G** and **B** are unity values for red, green and blue. These values give white light when combined. For given colour matching functions, any colour can be described as a set of tristimulus values R, G and B that express the necessary amount from each primary colour in order to perceive the reference colour at the left hand side of the above equation. When a different set of matching functions are available, the existing set of tristimulus values can be mathematically transformed to another set of values that uses different primary colours [SAM09]. Transformations between tristimulus values can be inverted and/or applied collectively depending on the application.

The matching functions that correspond to the RGB tristimulus values can also get negative values. As it is not feasible to add negative light values the amount of light that needs to be subtracted is simply added to the target colour i.e. the one that the three tristimulus values represent. The CIE developed a set of positive matching functions with a new set of associated tristimulus values namely, X,Y and Z. These tristimulus values are organised such that the Y channel represents the luminance of the target colour. It is worth noting that these tristimulus values correspond to primary colours that cannot be visualised as was the case with the primary colours that correspond to the RGB tristimulus values [Gue90].

Every set of tristimulus values defines a coordinate system where any colour can be expressed as a linear combination of these values. This set of basis vectors is known as *Colour space* while all the possible colours that can be produced by this basis is termed the *Colour gamut* of the colour space. It is common practice to project the three dimensional space that is spanned by the colour gamut into a 2D space known as *cromaticity diagram*. This projection is carried out by normalising every tristimulus value and discarding the luminance channel (Y in an XYZ colour space) [Wri81].

2.1.3 Visual display methods

This section summarises the imagery display technologies that are frequently used to display visual stimuli. The luminance range of the visual display is used for distinguishing between Low Dynamic Range (LDR) and High Dynamic Range (HDR) displays.

Dynamic range

Dynamic Range is the ratio of the lowest and highest measurable light intensities in an image or video. In real life the term can either be used to describe human's eye ability to see from very dark to very bright light conditions or it can be used to describe the level a display monitor (or a capture device) is capable of representing true black or white colours.

In Low Dynamic Range (LDR) every pixel contains an 8-bit value for every one of the Red, Green and Blue channels resulting to a 24-bit true colour information or approximately a range of 16.7 million colours that can be represented in a traditional display. In many cases 32-bits are reserved for a pixel and these are allocated as 24-bits true colour and 8-bit for alpha transparency. Typical LCD monitors have around 250-300 cd/m^2 while the luminance of an LCD television is usually between 300-600 cd/m^2. This quantity is usually higher for plasma displays and LED LCD TVs (around 1000 cd/m^2). Higher levels of luminance means that the device is capable of displaying more vivid colours as well as enhanced levels of brightness.

The human eye is a sensitive organ that can perceive a very large luminance range depending on observer's adaptation to the ambient conditions. Specifically, in very dark surroundings the human eye can perceive luminance levels as low as 0.001 cd/m^2 which is the lower limit of the Scotopic vision regime. This threshold has been quantified as the ability of an observer to see a candle at 30 miles on a clear night. It corresponds, on average, at around 90 photons falling at the cornea or 9 photons at the retina [Wal02]. The upper limit of the luminance range is approximately 10^4 cd/m^2 and is achieved under bright daylight conditions. This threshold defines the higher end of the Photopic vision regime where vision acuity and perception of colours are both at high levels.

Current display technology is capable of representing only two orders of magnitude of this dynamic range [SHS$^+$04]. On the other hand, the human eye is able to approximately perceive five orders of magnitude at the same time under normal lighting conditions. When abrupt transitions of the lighting conditions occur the

eye needs time to adapt to the new illumination conditions and the adaptation time depends on whether the transition is from a bright to a dark environment (dark adaptation) or the other way around (light adaptation) [AR96].

High dynamic range imagery takes into account a higher dynamic range aiming to approximate HVS's luminosity range. An HDR image is either composed of a series of LDR images at different exposures that contain subregions of the dynamic range or it results as a computer rendered image. As an HDR image contains a wider range of light information, its capture, representation to a display and storage follow more advanced techniques compared to those used in LDR images.

2.2 Introduction to sound

This section describes the main physical properties of sound as well as how humans perceive sound stimuli in an environment. Sound waves are created as pressure vibrations of a medium [Gle15] and can be digitally represented as amplitude values over time (waveform representation).

2.2.1 Physical properties of sound

In contrast to light, sound cannot travel through the vacuum and a medium (solid, gas or liquid) is necessary for the propagation of the sound waves. Through solids sound propagates as a transverse wave; that is a wave whose oscillations are perpendicular to the direction of propagation. In liquids and gases, sound propagates in the form of a longitudinal wave, these waves oscillate in the direction of propagation.

The speed of sound in a medium depends on its density and its elastic properties. The elasticity of a medium expresses medium's ability to retain its shape due to external forces that are applied upon it. The Newton-Laplace equation [BK15] can be used to calculate the speed of sound through a medium as:

$$c = \sqrt{\frac{K}{\rho}}, \tag{2.5}$$

where ρ is its density and K is the elastic bulk modulus of the medium. The bulk modulus expresses medium's ability to deform when pressure is applied upon it and its SI unit is the same as pressure (Pascals). Therefore, sound travels slower in mediums of low elasticity (e.g. rubber) but it increases when travelling through materials that cannot be deformed easily (e.g. wood) or have low density. These speed variations of the sound waves cause time delays that are audible to humans [Hel15].

In gases the speed of sound depends on the temperature. For example, the speed of sound in air under a range of temperatures (from −35 °C to 35 °C) can be calculated as:

$$c_{\text{air}} = 331.3\sqrt{1 + \frac{T}{273.15}} \tag{2.6}$$

where T is the temperature measured in Celsius degrees. At room temperature of 20 °C the above formula gives that the speed of sound in air is $c_{\text{air}} = 343.2$ m/s.

Joseph Fourier in 1827 demonstrated that any periodic waveform can be written as a series of sinusoidal waves of different frequencies which are integer multiples of a fundamental frequency [Fou27]. For example, the middle **C** (261.6 Hz) note played in the guitar can be written as a sum of sinusoids at frequencies of 261.6 Hz, 523.2 Hz, 784.8 Hz and so on. A sinusoidal wave is a simpler way to represent a sound wave and can be written in the form:

$$f(t) = A\cos(\omega t + \phi), \tag{2.7}$$

where A is the amplitude of the wave, ω is its angular frequency measured in rad/sec and ϕ is its phase measured in radians. The angular frequency is given as $\omega = \frac{2\pi}{T} = 2\pi f$ where T is the time of one period and f is the frequency of the sinusoid in Hz. As was the case with light, the frequency and speed of a sound wave are connected as:

$$\lambda = \frac{c}{f}, \tag{2.8}$$

where λ denotes the wavelength of the sinusoid. Image 2.1 shows pictorially the physical quantities of a sinusoidal wave.

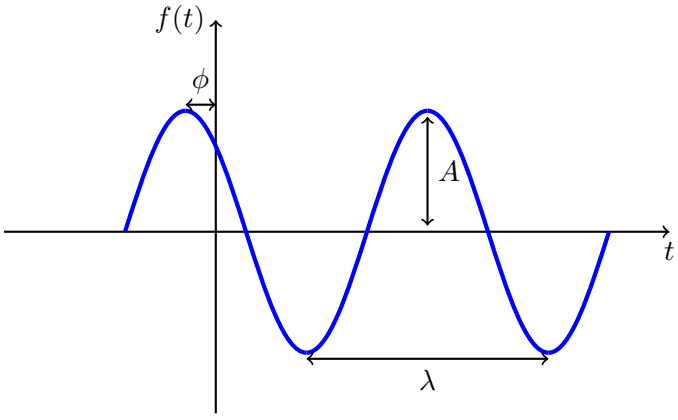


Figure 2.1: Physical properties of a sinusoidal wave.

At standard temperature and pressure conditions ($T = 20$ °C and $P = 101.325$ kPa [NIS01]) the range of audible wavelengths starts from 0.01715m up to 17.15m. In terms of frequency the audible spectrum is between 20 Hz to 20 kHz and it is usually divided into smaller subregions known as *Octave bands*. Every octave band contains a centre frequency that is the representative of the band and is used to calculate the lower and upper bounds of the band. The seventh octave band centre frequency is defined as $f_7 = 1000$ Hz and can be used to derive the centre frequencies of the other bands using the formula: $f_{n-1} = \frac{f_n}{2}$ for $n = 2, \cdots, 11$. The lower and upper bound of each band can be given by the formulae:

$$f_n^{\text{low}} = \frac{f_n}{\sqrt{2}} \text{ and } f_n^{\text{high}} = \sqrt{2} f_n. \tag{2.9}$$

The bandwidth of each octave band is constant and approximately equal to 70.7%.

Sound intensity and pressure

Sound intensity, denoted as I, describes the sound power per unit area. The power that is released from a sound source or received to a surface can be measured in Watts, therefore, sound intensity is measured in Watt/m^2. Sound pressure, denoted as p, expresses the oscillation of a medium's molecules due to the effect of a sound wave on them. Sound pressure can be measured at any point using microphones which contain sensitive diaphragms that oscillate when triggered by the incoming sound wave. The microphone samples the sound wave and a digital signal that represents the waveform of the sound is created. Sound pressure is measured in N/m^2 or Pascals.

Sound intensity and pressure are related to each other and specifically intensity is proportional to the square of the pressure [BS04]:

$$I \propto p^2. \tag{2.10}$$

Sound intensity and pressure both fade according to the inverse distance law. The inverse distance law can be stated in terms of pressure at two different measurement points as:

$$\frac{p_2}{p_1} = \frac{r_1}{r_2}, \tag{2.11}$$

where r_1, r_2 are the distances of two points away from the sound source and p_1, p_2 are their corresponding pressures.

Acoustic attenuation occurs when the sound waves interact with the surfaces of the environment and depends on the frequency of the sound and the surface

material. When a sound wave interacts with the environment a portion of the acoustic energy is transmitted into the surface while the rest of it is absorbed or reflected back in the environment. The amount of absorbed acoustic energy depends on the absorption coefficient $a(f)$ of the material and expresses the ratio between the absorbed and incident acoustic energies. The value $a(f)$ depends on the frequency of the impingent sound and it ranges between 0 and 1 where the value 1 is attributed to a material that absorbs all the incident acoustic energy while 0 indicates that all the incident acoustic energy is reflected. Tabulated absorption coefficient data can be found for various materials and across all the frequency octave bands.

2.2.2 Perceptual Properties of sound

The hearing dynamic range, in terms of sound intensity units, starts from $I_0 = 10^{-12}$ W/m^2 up to $I_{\max} = 10$ W/m^2. I_0 is the lowest intensity of a pure tone at 1000 Hz that can be perceived by a human listener and is known as the threshold of hearing. $I_{\max}$ is the highest audio intensity that can be tolerated by a human listener and is known as the threshold of pain. This vast range can be compactly measured using logarithmic values of the ratio $\frac{I}{I_0}$ as follows:

$$I(\text{dB}) = 10 \log_{10} \left(\frac{I}{I_0} \right), \tag{2.12}$$

where dB is the shorthand notation for *decibels*. The advantage of measuring sound intensity relative to the threshold of hearing I_0 is that the wide range of sound intensities can be measured as the range of 0 to 130 dB where 0 dB is the threshold of hearing and 130 dB is the threshold of pain. This transform is intuitively related to the way human perception works as the perceived intensity of a stimulus changes logarithmically when its physical magnitude increases.

The same relative measurement method can also be used for sound pressure. In that case, the threshold of hearing can be written as $p_0 = 2 \cdot 10^{-5}$ Newton/m^2 and any other pressure level can be calculated in dB as:

$$p(dB) = 10 \log_{10} \left(\frac{p^2}{p_0^2} \right) = 20 \log_{10} \left(\frac{p}{p_0} \right), \tag{2.13}$$

Equations 2.12 and 2.13 are equivalent as the sound intensity is proportional to the square of the pressure (see formula 2.10).

Sound intensity and frequency are physical quantities that are closely related to the perceptual quantities of *Loudness* and *Pitch* respectively. Loudness, as the name suggests, expresses how "loud" or "soft" a sound is perceived and is connected

to sound intensity via Steven's law:

$$L = kI^a, \tag{2.14}$$

where L is the loudness measured in sones, I is the sound intensity, k is a proportionality constant that depends on the measurement units and a is a parameter that depends on the sensory stimulus measured (e.g. loudness, brightness, contrast, taste, etc.). One sone is the loudness unit for a 1000 Hz sound at intensity of 40 dB. Pitch is connected to the fundamental frequency of a tone (see section 2.2.1) and expresses how bass or tremble a sound is perceived. Pitch is measured in cents and there are 1200 cents in every octave band.

Figure 2.2 (left) shows how the JND threshold for frequency scales for sounds of low and medium frequencies. At low frequencies the average human listener can easily perceive a pair of sounds as different as the JND threshold is low (approximately 1 Hz). At higher frequencies the JND threshold increases and perceptually distinguishing pairs of sounds becomes more difficult. For sound intensity the JND threshold depends on the frequency of the tone and the intensity level. For example, a 60 dB sound at 200 Hz has a JND threshold of around 0.8 dB. This values change for different frequency sounds. Figure 2.2 (right) shows how the JND threshold scales for different frequencies and sound intensity levels.

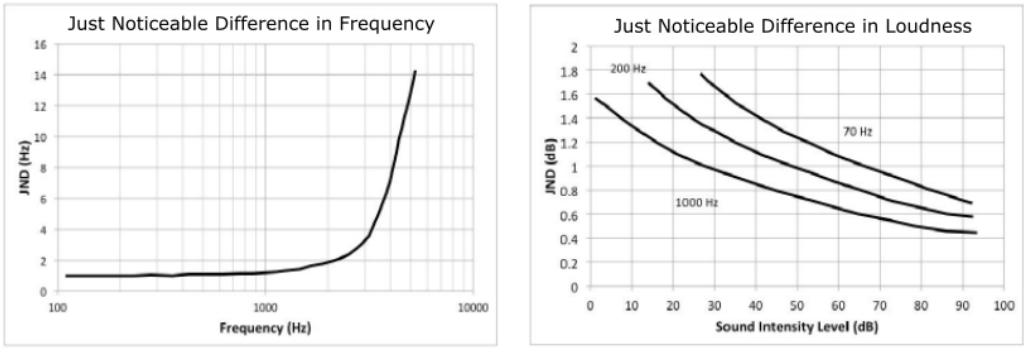


Figure 2.2: **Left:** Just Noticeable Difference for frequency. **Right:** Just Noticeable Difference for sound intensity level at frequencies 70Hz, 200Hz and 1000Hz. Material is adopted from [For13] with author's permission.

2.2.3 Sound localisation

Humans have the ability to detect the direction of an incoming sound in an environment and the relative distance of the sound source away from them [Bla96]. As

opposed to the eyes that can capture only a portion of the environment, the human ear is able to collect auditory information from all directions using *binaural* and *monaural* cues. Binaural cues can be either the Inter-aural Time Difference (ITD) or the Inter-aural Level Difference (ILD) between the two ears. ITD is caused due to the extra time needed for the sound wave to travel around the head and arrive at the ear that is physically further from the sound source. ILD, on the other hand, is the result of a phenomenon known as the *Shadowing Effect* [VWVO04]. The human head obstructs a portion of the original acoustic energy so the sound wave has lower amplitude when arriving at the distant ear. The two ears also receive individual monaural cues that contain information on how the sound waveform is affected by the head shape, shoulders, torso and/or other anatomic characteristics before entering the ear.

Inter-aural time difference

Lord Rayleigh's Duplex theory [Ray07] was the first study which investigated the problem of sound localisation. This theory assumes that the human head is a completely spherical surface and the wave will reach one of the two ears faster than the other depending on the angle of the incoming sound (see Figure 2.3). The ITD is defined as:

$$\text{ITD} = \frac{r}{c}(\theta + \sin\theta), \quad -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}, \tag{2.15}$$

where c is the speed of sound in air, r is the radius of the head and θ is the azimuth angle of the incoming sound wave. If the sound source is directly ahead, ITD $= 0$ and there is no time delay while if the source is off to one side then $\text{ITD}_{(\max)} = \frac{a}{c}(\frac{\pi}{2} + 1)$ and a time delay of about 0.7 ms is caused for a human head of typical size.

Inter-aural level difference

Duplex theory explains that the bending of the sound wave around the head is also responsible for amplitude differences between the two ears. The ILD has been found to depend not only on the azimuth angle but also on the frequency of the incoming sound wave. At low frequencies the wavelength is long compared to the head diameter and the sound pressure that is exerted at the two ears is approximately the same. At high frequencies the wavelength is quite small and the difference in sound pressure between the two ears can be as high as 25 dB. This difference causes the head shadow effect.

According to the Duplex theory, the ITD and ILD operate at different frequency ranges. Specifically, inter-aural timing cues exist at all frequencies but for

periodic sounds they can be decoded for frequencies lower than 1.5 kHz [Ray07]. At low frequencies the sound wave bends around the head for less than a full circle (2π radians) and the ITD gives meaningful azimuth directional information. At frequencies higher than 1.5 kHz where the head size is comparable to the wavelength of sound the ILD dominates in the localisation perception as the ITD cannot provide meaningful directional information due to the multiple bendings of the sound wave around the head. Duplex theory claims that ITD and ILD can provide sufficient sound localisation information for the whole audible spectrum.

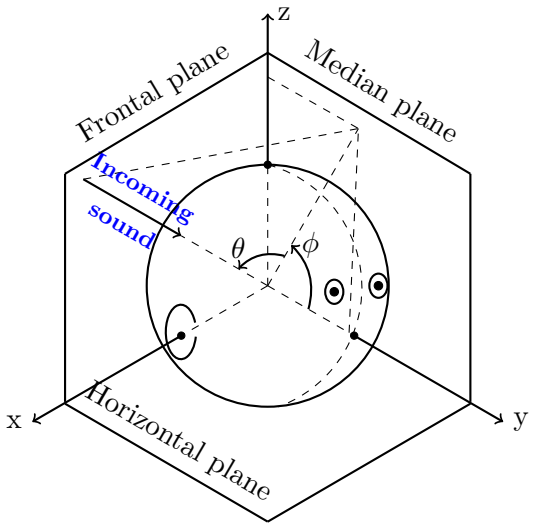


Figure 2.3: Sound waves reach the head at an azimuth angle θ and an elevation angle ϕ. The image also depicts the Median, Horizontal and Frontal plane that are used for sound localisation. The definitions given in the image are described in the CIPIC HRTF database [ADTA01].

Head related transfer functions

The use of ITD and ILD can provide accurate localisation information when sound is coming at an azimuth angle on the left or right of the listener. In case that the sound signal arrives at an elevation angle either above or below the head (see Figure 2.3), the use of time or level differences leads to an infinite number of points which have the same ITD or ILD. These points are distributed in a cone that extends outwards from the ear and is known as the *Cone of confusion* [Bla96].

When a sound waveform approaches the human head it is reflected off the surfaces that compose the head, shoulders, pinna (outer ear) before entering the auditory canal. These reflections slightly modify the shape and physical character-

istics of the original waveform creating what are known as monaural cues. Monaural cues can be captured and encoded in the form of digital signals as described in the following.

The monaural cues can be captured for the left and right ear in the form of two signals $h_L(t)$ and $h_R(t)$ that are termed Head Related Impulse Responses (HRIR). The Fourier transform of each one of these signals is known as Head Related Transfer Function (HRTF) and encodes localisation information between the sound source and each one of the two ears. Figure 2.4 shows a graphical representation of an HRIR in time domain at the horizontal and median planes (these planes are given pictorially in Figure 2.3). The convolution of a sound $x(t)$ that emanates from a sound source with each one of the HRIRs yields signals $x_L(t)$ and $x_R(t)$ that contain acoustic localisation information for the left and right ear respectively. They can be calculated as follows:

$$\begin{aligned} x_L(t) &= h_L(t) \otimes x(t) = \int_0^\infty x(s)h_L(t-s)ds \text{ , for left ear.} \\ x_R(t) &= h_R(t) \otimes x(t) = \int_0^\infty x(s)h_R(t-s)ds \text{ , for right ear.} \end{aligned} \tag{2.16}$$

The above methodology gives localisation information for both elevation and azimuth but assumes that the sound wave does not interact with any other objects or surfaces between the source and the receiver. In real applications sound waves interact with the surfaces of the environment yielding more localisation cues (e.g. reverberation effects). The effect of these cues can be taken into account and added to the source signal by simulating the basic acoustic features of a room before convolving with the HRIRs. The omission of these sound interactions leads to sounds that are perceived unusually close or inside listener's head meaning that there is no sense of sound externalisation.

Phase dependence

When two sinusoidal waves of the same frequency have the same amplitude and propagation speed (both magnitude and direction), they can be combined either constructively or destructively depending on their phases. This physical phenomenon is known as wave *superposition*. Specifically, two sinusoidal waves that have the same amplitude A and their phases are ϕ_1 and ϕ_2 respectively are considered *in-phase* when $\phi_1 - \phi_2 = 0$ for $0 \leq \phi_1, \phi_2 < 2\pi$. In this case the resulting wave sounds louder as the amplitude is doubled. On the other hand, when $\phi_1 - \phi_2 = \pi$ then the two

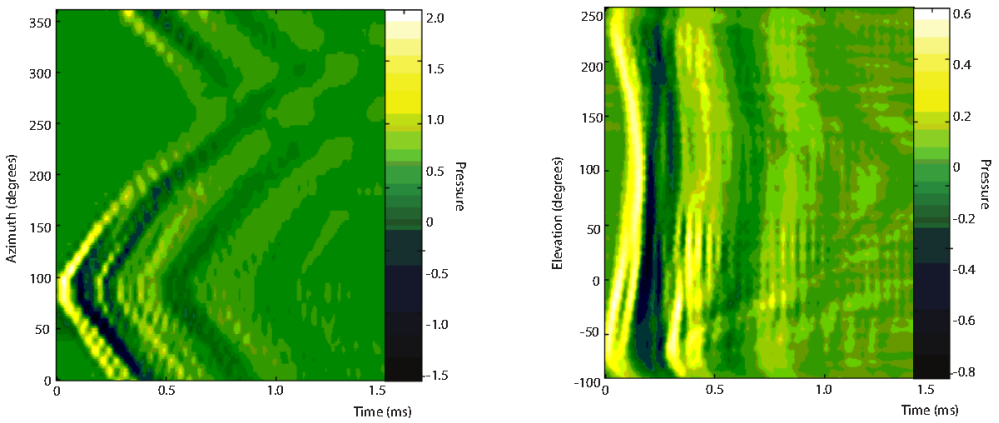


Figure 2.4: Example data from the KEMAR HRTF database [GM94]. **Left:** Pressure that is exerted on the right ear due to a sound source that starts from ahead ($\theta = 0°$) and rotates at the horizontal plane. Highest pressure values occur when sound is coming directly from the right ($\theta = 90°$) and are depicted with bright colour. When sound comes from the left side ($\theta = 270°$) the pressure level measured at the right ear is the lowest. **Right:** Pressure values measured in the median plane. As the sound source moves on the horizontal plane, pressure values do not show the same pattern as previously instead the sound arrival times remain relatively unaffected.

waves interfere destructively (*out-of-phase* waves) and the result is a waveform of 0 amplitude. An example that wave superposition can be frequently noticed is in cinemas or large auditoriums. In this places, sound originates from many different sources and superposition might lead to silent spots for out-of phase waves or it can act additively, leading to places where sound is unusually loud (*sweet spots*).

2.2.4 Auditory display methods

This section describes methods to estimate HRTFs when presenting sound to a human listener. The description is not only limited to specific HRTF estimation methods but also extends to techniques that are used to provide aural stimuli enriched with sound localisation cues. These techniques are well fitted to particular auditory display devices as it is described in the following.

HRTF approximation methods

The calculation of HRTFs is highly dependent on the anatomical characteristics of the listener thus the derivation of an exact transfer function for every individual is not a trivial task. CIPIC [ADTA01] has conducted HRTF measurements in an anechoic environment for many human subjects with different head anatomical characteristics that represent different groups of the population. Given the anatomic characteristics of a human listener, someone can choose the best HRTF from the database that matches the specified criteria.

Another way to obtain high quality HRTFs is to measure individual's anatomic characteristics using head anthropometry measurements as proposed by CIPIC. This method includes various measurements of the pinna (outer ear), head, neck, torso and shoulders and produces an HRTF tailored to the individual that is being measured. This procedure requires a lot of time so it is very difficult to be systematically used by commercial sound systems even though it gives the best possible acoustic experience for the listener being measured.

Many studies [LNL06, SNH$^+$10] propose models that can estimate an HRTF to a certain accuracy. These models include parameters that account for the anatomical characteristics of the listener and despite their known limitations give satisfactory solutions with little computational cost. A simple HRTF model that can be used for sound localisation in the horizontal plane is the ITD model. This model introduces a time delay at the sound signal that arrives at the two ears depending on the azimuth angle of the incoming wave. The time delay can be calculated as:

$$T_D(\theta) = \begin{cases} \frac{a}{c}(1 - \cos\theta) & \text{if, } |\theta| \leq \frac{\pi}{2} \\ \frac{a}{c}(|\theta| + \frac{2-\pi}{2}) & \text{if, } \frac{\pi}{2} \leq |\theta| \leq \pi, \end{cases} \tag{2.17}$$

where a is the radius of the head and c is the speed of sound. This model can be implemented in the form of a Finite Impulse Response (FIR) filter and applied to the source signal through convolution. Its application does not provide sound externalisation and elevation cues.

Another HRTF model that can be used for localisation cues in the horizontal plane is the ILD model. This can be implemented as a realizable Infinite Impulse Response (IIR) filter using the following transfer function:

$$H(s, \theta) = \frac{f(\theta)s + \beta}{s + \beta} \tag{2.18}$$

where $f(\theta) = 1 + \cos\theta$ and $\beta = \frac{2c}{\alpha}$. In this equation s denotes the sound pressure level in dB and a, c are head radius and speed of sound respectively. This is a

transfer function with one pole (root of the denominator) and one zero (root of the nominator) and it enhances the high frequencies when the azimuth angle θ is 0 degrees while it fades them when θ is 180 degrees which is exactly what happens at the head shadowing effect. Similarly to the ITD model, the ILD model cannot simulate the effect of source externalisation and it is not able to provide elevation cues.

Another way to approximate an HRTF is through the combined use of both ITD and ILD models [BR97] and switching between the two depending on the frequency of the incoming sound. This technique overcomes the ambiguity of the ITD model to resolve localisation information for sounds of high frequency. This model still provides localisation information only at the horizontal plane without resolving the sound source externalisation problem.

Brown and Duda proposed an HRTF model [BR97] that includes the combined use of the ITD and ILD (formulae 2.17 and 2.18) as a general head model that captures the interaction of the incident sound with listener's head. They complement the head model with pinnae and shoulder models so as to produce a pair of left and right HRIRs that can capture both elevation (ϕ) and azimuth (θ) localisation effects. The authors resolve the problem of externalisation by introducing a room echo model which produces room reverberation effects by introducing delays to the signal that is received from the head model. The proposed model can adapt for listeners of different anatomic characteristics through the use of different parameters for every one of the body models.

Binaural Techniques

Binaural techniques are methods for displaying 3D sound to a listener using headphones [JLW95, Møl89, Møl92]. These methods provide a more immersed acoustic experience compared to the traditional mono and stereo techniques that can provide limited directional information. In binaural techniques every ear receives a different audio stream which has specifically been prepared for it. Every incident sound signal $s_I(t)$ that originates from a sound source at location (r, θ, ϕ) is convolved with a pair of HRIRs $h_L(r, \theta, \phi, t)$ and $h_R(r, \theta, \phi, t)$ that can be either measured beforehand or obtained from a database that contain a range of different impulse responses (see section 2.2.3). The output signals, known as binaural responses, contain spatialisation cues that are specifically designed for the listener whose anatomic characteristics are encoded in the left and right HRIRs. The process of calculating binaural responses is repeated for different locations of the sound source or the receiver so as the resulting audio stream to contain accurate spatial information. It

is worth mentioning that if the original sound stream does not contain the effects of the room (i.e. it is an anechoic sound stream), these can be added to it through the use of a Room Impulse Response (RIR) that contains reverberation information specific for an environment. The design and properties of a room impulse response are explained later in section 6.2.

Except from using headphones one can deliver binaural cues through the use of loudspeakers located in front of a listener. This can be done by stopping the sound signal that emanates from the right (or left) loudspeaker to reach the left (or right) ear and is known as *crosstalk cancellation* [ACB+07]. This method can be easily applied to a stereo or multi-channel system through the use of Digital Signal Processing techniques (DSP) [RM15].

Binaural techniques are also used for efficient recording of audio in an environment. In that case a head dummy with microphones at each ear is used to capture monaural and binaural localisation cues as a real human head would do.

Monophonic, Stereophonic and Multi-channel Techniques

A simple method to display sound to a listener is through the use of one (Monophonic) or two (Stereophonic) audio streams that are delivered using loudspeakers or headphones [Ear89, SE98]. Monophonic sound, as the name indicates, is used for displaying sound that originates from one direction. Stereophonic or stereo are techniques imitate the way ITD and ILD localisation methods work. For a pair of loudspeakers located in front of the listener at the horizontal plane, different delay and gain levels are applied for each of the two signals that are directed to the two loudspeakers. This gives the impression that sound is coming from different directions in the horizontal plane.

A physical extension to stereophonic techniques it to include multiple audio channels for displaying audio from many different loudspeakers arranged around a listener in the horizontal plane [DFMM99]. This method is known as surround sound and sometimes includes loudspeakers arranged in space at different azimuth and elevation angles around the listening area. Popular sound surround configurations are the 5.1 or 7.1 systems [CCW03]. The former includes three loudspeakers in front of the listener and two at the rear for a total of 5 channels. An audio channel is also used for displaying low frequency bass sounds (subwoofer). The 7.1 loudspeaker configuration is the same as 5.1 with two additional audio channels normally located at the left and right of the listener (plus a subwofer). These sound surround systems are popular choices for home entertainment, video games or other interactive applications.

2.3 Wave phenomena

Sound and light are transported in the form of waves. These waves interact with the surfaces of the environment resulting to wave phenomena such us reflection, refraction, diffraction (see Figure 2.5) and interference.

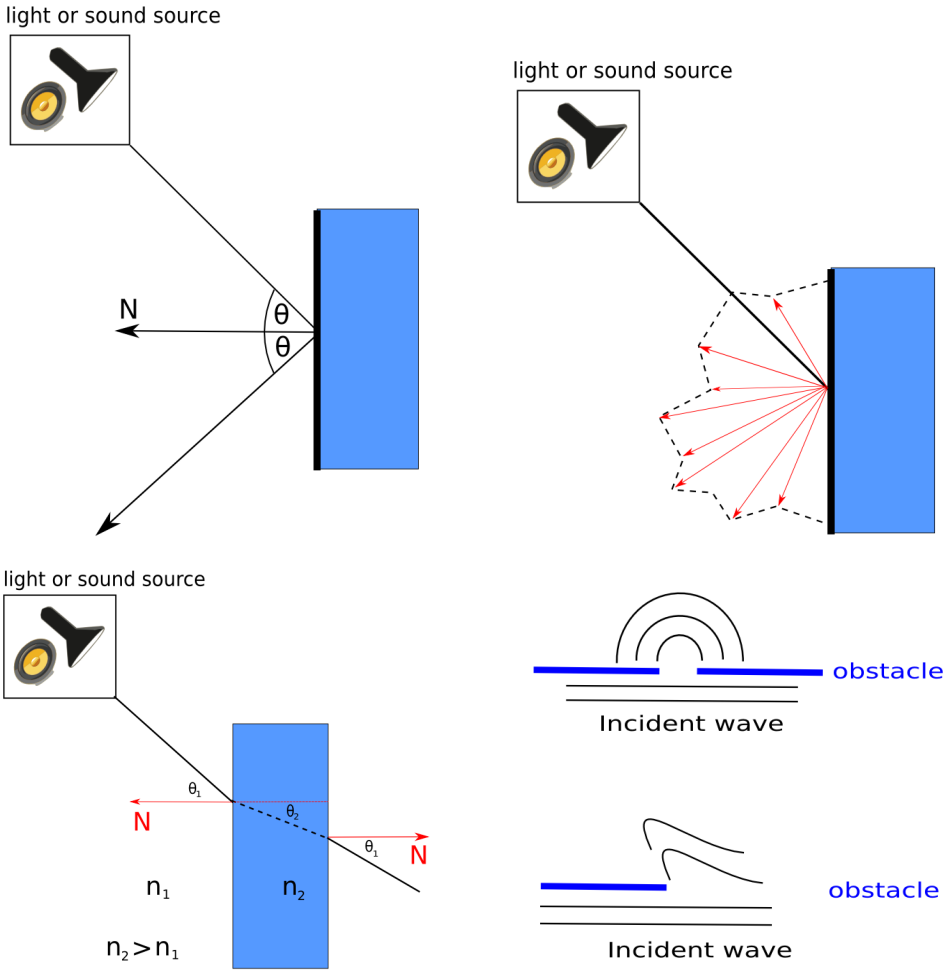


Figure 2.5: Light and Audio wave phenomena. **Top-Left:** Reflection, **Top-Right:** Scattering, **Bottom-Left:** Refraction, **Bottom-Right:** Diffraction. Images of the physical phenomena were adopted by [HRW13]

Reflection

The law of reflection describes how an incident wave of angle θ can be reflected and return back into the medium from which it had originated before the interaction with the surface. This reflection can be specular where the wave is reflected with

the same angle θ or it can be diffuse depending on the reflectivity properties and roughness of the incident surface. Reflection also depends on the incident angle and the direction of oscillation of the incoming wave (longitudinal or transverse waves). For light waves, reflection properties can be estimated for various materials using Fresnel's equations [HZZ01].

Refraction

Refraction is the phenomenon where the direction of an incident wave changes when entering a different mediumwith different refraction index. The speed of light in a medium depends on its refraction index and when light enters from a medium of low refraction index (e.g. air) to a medium of higher refraction index (e.g. water) it bends towards the surface's normal vector. If the two media have refraction indices n_1 and n_2, Snell's law can be used to calculate the angle that the wave propagates into the medium with the higher refraction index. This is given as:

$$\frac{n_1}{n_2} = \frac{\sin\theta_1}{\sin\theta_2} \tag{2.19}$$

where θ_1 is the incident angle and θ_2 the angle of the refracted wave with the normal vector N when propagating into the medium with refractive index n_2 (see Figure 2.5). When light refracts, the wavelength of the incident wave either increases or decreases but its frequency remains constant. This causes the speed of the wave to change accordingly depending on the media that the wave propagates into.

Diffraction

Sound and light waves can propagate through openings or around barriers that are found on their path. This phenomenon is known as diffraction. The effects of this phenomenon are more noticeable when the wavelength of the propagating wave is comparable to the geometry size of the object that the wave diffracts upon (see also Figure 2.5).

Interference

Interference of waves refers to the phenomenon where two waves are superimposed either constructively (in-phase) or destructively (out-of-phase) and was explained in section 2.2.3.

2.4 Introduction to olfaction

Human perception is based on stimulations coming from multiple senses, namely, vision, hearing, olfactory, gustatory and tactile sensory stimuli. The significance of the sense of smell, in particular, is not initially obvious but it affects how someone thinks, feels and behaves [SP04, PSNG06]. The sense of olfaction is related to recalling memories associated with a perceived scent, a phenomenon known as *olfactory-evoked memory*. This phenomenon is attributed due to the direct link of the olfactory cortex with the limbic system in the human brain. The limbric system is responsible for recalling previous memories or adjusting our emotional state [HE96].

The sense of olfaction has not been given the same merit compared to the other senses that are frequently used in Virtual Reality (VR) applications. This can be mainly explained due to the limited knowledge of the human olfactory system and the way it receives and processes odour stimuli. The human nose can discriminate more than 1 trillion smells using an underlying mechanism that is composed of multiple and diverse olfactory receptors [BMVK14]. The majority of these smells are composed of many different chemical compounds that cannot be discriminated individually. When their mixture is received by the nose the Human Olfactory System (HOS) perceives them as a single odour. This phenomenon is known as *perceptual blending* [LL98] and is quite significant in the sense of olfaction.

Smell perception contains three basic stages which are: estimate the intensity of the smell, qualitatively describe the perceived smell and finally express subjective opinion of how pleasant or not is the perceived smell (hedonic tone) [HCL03]. Smell detection and identification is highly subjective although there are smells that can be identified in concentrations as low as 4×10^{-15} grams/litre and recognised in concentrations of 2×10^{-13} grams/litre. These measurements represent statistical estimations of the population and are subject to human variability [Nef98, PSNG06].

This section reviews the physical characteristics of odorous molecules and explains the two physical processes that govern smell transport. Following this, a description is given on the how smells are assessed and classified by humans. Finally, the state of the art technology for displaying and analysing odours is presented.

2.4.1 Physical properties of smell

Volatile organic compounds

Volatile Organic Compounds (VOC) compose the majority of the odours we perceive [BMVK14]. These are carbon-based chemical substances that have high vapour

pressure at ordinary room temperature conditions. These compounds have a low boiling point which is the reason why large numbers of molecules can evaporate from their liquid or solid form into the ambient air over a period of time (*volatility property*). There is a large number of VOCs that are inhaled by the human olfactory system on a daily basis and many of them are odourless, while others can be toxic when inhaled in large amounts. Health and safety standards report lists of VOCs that can be harmful when received in large concentrations. These standards also propose safe ranges of temperature and pressure in different places (working environments, hospitals, schools, etc.) as a measure of precaution for preventing faster evaporation rates of these compounds in the ambient air [CEN03].

Odour concentration

When an odour is spread in the environment, it mixes with the air and/or other odours creating a homogeneous mixture (smell-air mixture) [ADP01]. The amount of the odorous substance characterises how diluted or concentrated is the smell in the mixture and can be measured using parts per million (ppm), parts per billion (ppb) (1 ppm = 10^3 ppb) or parts per trillion (ppt) notation (1 ppm = 10^6 ppt). This is a relative measure that is frequently used to describe two dimensionless quantities known as *mass fraction* and *molar fraction*. Mass fraction is given as:

$$w = \frac{m}{m_{\mathrm{T}}} \tag{2.20}$$

where m is the mass of the odour and m_{T} is the mass of the mixture. In mole fraction the above ratio is expressed in terms of *Moles* for both the odour and the mixture instead of masses. Both of these quantities can be used to quantify the amount of odour relative to the amount of the total mixture. For example, 1 ppm of a specific odour (e.g. aroma) means that 1 kg of mixture will contain 1 mg of odour. Measurements in ppm or ppb are practical as they allow to describe concentrations of tiny amounts of odour in large air mixtures.

Diffusion and convection

The motion of odorous particles can be described through a physical process termed *Diffusion*. This process describes the random motion of the particles from areas of higher concentration to areas of lower concentration. The diffusion of odour particles into air is a very slow process and according to Dusenbury [Dus92], a particle that diffuses for one minute can, on average, cover a distance of only 70 mm in an environment with no external perturbations (e.g. air draughts, temperature

or pressure gradients, etc.). The phenomenon of diffusion is not strictly limited to describe odours movement but it can be generally used to describe the spread of one substance into another when the two substances are either liquids or gases. Similar is the concept of energy diffusion in the form of heat. Heat disperses from areas of higher temperature to cold regions of a room.

Apart from diffusion there is one more natural process that accounts for smell transport in a medium and is termed *Convection*. Convection is the process that describes the bulk transport of fluid mass due to the flow of the fluid in an environment. For example the flow of an air draught through a window can convect a smell plume in a room. In that case the spread of the smell in the environment will be achieved much faster than in the case where the odour is transported in a sealed room through diffusion.

Convection in an environment can be caused due to different temperatures or pressures at different areas of the domain. Using the same example as before, someone can imagine a closed room with a hot radiator in it. The temperature gradient between the hot radiator surface and the cold environment causes large amounts of air to circulate around the room (heat transfer) and disperse a smell plume. Another example, is water flow through a pipe. The application of different pressures at the two ends of the pipe causes water movement in the direction from higher to lower pressure. The physical phenomena of convection and diffusion are sometimes collectively termed as *Advection* depending on the physical context of the problem under consideration.

2.4.2 Perceptual properties of smell

Smell habituation and anosmia

This section describes fundamental properties of the Human Olfactory System (HOS) when interacting with olfactory stimuli at various concentrations. Temporal attenuation of olfactory perception is a phenomenon that happens in a daily basis even though we might not realise it. Permanent loss of olfactory perception is rare and can be caused due to a damage to the olfactory bulb and/or the brain.

The HOS is able to adapt to the concentration levels of an odour present in an environment over time. The phenomenon of smell habituation or smell fatigue is the temporal inability to distinguish an odour after a prolonged exposure to this smell at concentrations that can be easily perceived. When an intense smell is present for a long time in an environment the HOS temporarily suspends the signals that are sent to the brain in an attempt to protect it [SP04]. Exposure to strong

smells can cause headaches that can be partially prevented by olfactory adaptation. Smell sensitivity is restored over time when the smell intensity fades or it is masked by another smell that needs to get identified and perceived by the HOS. When the concentration of the smell fluctuates over time the HOS adapts slower to it depending on the fluctuation level and the exposure time [CMA+10].

The loss of smell, known as *Anosmia*, is different than smell adaptation as it refers to someone's permanent inability to smell. According to the National Health Service (NHS) in UK around 6000 people were affected by anosmia at 2012 [HSN12]. Another form of anosmia is the inability to recognise a particular smell or family of smells which is known as *Parosmia*. Also there are cases of people that suffer from *Phantosmia* where they smell an imaginary smell that does not actually exist in the environment. Loss of smell can happen due to simple causes such as colds or obstruction of the nostrils or it can be the result of more rare symptoms like brain and/or nervous system damages.

Odour assessment

Smell assessment refers to the process of evaluating an odour using human perception criteria and not physical measurement units. Odour assessment is a procedure undertaken in the context of health and safety precautions at various environments were people live, work or get entertained [CEN03]. According to Nicell [Nic09] there are five criteria that are used for the perceptual assessment of an odour. These are: frequency, intensity, duration, offensiveness and location of a smell stimulus.

How frequently someone is exposed to smell stimuli defines the frequency of the presented odour and depends on the way smell is presented to him. For example, smell of constant concentration or bursts of smell. Intensity of smell refers to how strong a smell is perceived. Intensity can be considered as the perceptual analogue of concentration and can be estimated using Steven's power law (see equation 2.14). Writing this law in terms of smell intensity and concentration gives:

$$\log I = \log k + \alpha \log C, \tag{2.21}$$

where I is the perceived intensity, C is the concentration, and k, α are constants specific to the odour. According to Jiang *et al.* [JCT95] odour intensity can be divided to a scale of seven levels ranging from "no-odour" (level 0) up to "intolerable" odour (level 6).

Duration of smell refers to the time needed until humans adapt, on average, to the presented smell and again depends on the presentation method. Offensiveness

of smell is a subjective evaluation of how offensive an odour is. This is connected to the hedonic assessment of the presented smell and varies in a person to person basis although there are smells that are perceived as pleasant by the majority of people (e.g. chocolate smell) and others that are characterised as distasteful or unpleasant (hydrogen sulphide or the smell of rotten egg). Finally, location refers to the source of smell. Health standards set strict restrictions for the distance of unpleasant smell sources near work places, residential areas or leisure facilities.

Odour classification

Many studies have proposed classifications of odours into different groups depending on their perceptual impact on panels of human subjects. This process is highly controversial for many reasons. The sense of olfaction is connected to previous memories and experiences thus a classification is mostly based on subjective criteria rather than objective techniques [HCL03]. In addition, people from different demographic groups of the population might not even be familiar with the presented smell *per se* as they were never exposed to it previously. Amongst other factors that affect smell classification is the intensity of the presented smell, the way of presenting smells (continuous flow smells or puffs) as well as the impact of the surrounding environment when presenting a smell stimulus [WOC00]. An odour may not have the same perceptual impact to someone when is delivered in an experimentation room instead of the environment that is usually sensed. In the following, a brief overview of different classification schemes is presented. For more information the interested reader is referred to [HEL68, RSD$^+$02].

Ammore [Ern82] has proposed an odour classification into 7 odours that are frequently found in nature amongst a range of 600 organic compounds. These are: camphor, musk, floral, peppermint, ether, pungent, and putrid. Another scheme, as proposed by Rimmel [Rim10], considers a wider set of 18 main smells while in the same way, Zwaardermaker [Zwa95] proposed a more technical classification scheme that comprises 9 main smell categories that are subdivided into many subcategories. In another study by Henning [Hen16], the author proposes a classification scheme similar to the way the monochromatic RGB values are used in the description of colour perception. This classification scheme is composed of an odour prism where every smell can be given as a mixture of 6 main components in appropriate concentrations. The main components are: flowery, putrid, fruity, spicy, burnt and resinous. In a more recent study, Weiss *et al.* [WSY$^+$12] report evidence that the HOS can perceive a "white" smell in the same way as the HVS perceives the white colour as a neutral colour that results from mixing the highest possible values of

Red, Green and Blue. In their study, participants were asked to identify the components of an odour mixture where every component co-existed in the mixture at equal concentration proportions. The results indicated that at a mixture of around 30 smells, participants were no longer able to distinguish the different components. The authors argue that the prerequisites to obtain the olfactory "white" are firstly, equal proportions of the smells that compose the mixture and secondly, odours that can be easily perceived as different from each other.

Recent attempts to classify odours are based on the use of verbal descriptions for smells [Ste01]. This process assumes that all the participants have been presented with a particular set of smells in the past and are able to make use of specialised vocabulary in order to describe the displayed odour.

2.4.3 Olfactory display methods

Many studies have investigated the importance of delivering smell stimulations in virtual environments along with other senses. The introduction of realistic smell impulses can significantly increase user's immersion and sense of presence in VEs [Che06] while it can be used to enhance the level of realism in entertainment applications [YN06].

As opposed to the senses of vision and hearing, olfactory stimuli are quite momentary in nature and their delivery significantly differs from other senses. While visual and auditory stimuli can be precisely delivered as many times as necessary and evoke the same perceptual response, presentation of olfactory stimulations might be perceived differently at different delivery times. Also keeping a high level of temporal and spatial accuracy in vision and auditory is relatively easy while adjusting smell concentration over time and deliver accurate smell doses can be a challenging task.

The hardware that is used to present olfactory stimulations is known as *olfactory display* or *olfactometer*. These devices are designed with the objective to deliver precise amounts of smell through controlling the flow rate of the presented odour. They are frequently used for the implementation of various experimental studies in psychophysics and neuroscience [LGA+10]. In VR, the use of olfactory displays is relatively recent and devices designed for these applications are not yet fully customisable with existing VR technology [BD95]. In many cases, these devices are heavy and require a lot of space as many components need to be connected together such as smell reservoirs, pressure valves, mass flow controllers, air compressors, solenoids, air filters, etc.

State of the art olfactometers deliver smell stimuli that are produced as a mixture of many different odorous components in liquid form. These chemical

components are stored inside or near the device. The number of possible delivered smells is limited by the number of different odorous components and the possible mixtures that can be produced from them. Usually, the number of components ranges between one to six chemical substances. Mixing the chemical components in proportions that give meaningful odorous compositions is also another challenge that olfactometer manufacturers face. This is an evolved process and unless someone has expertise in perfumery, it cannot be easily settled [Mat10].

Designers take into account several features when constructing olfactory displays. These are computer controlled devices that permit the existence of multiple odour channels and one or more control channels. Control channels are used to dilute the smell that is coming out from odour channels with air or flush out the parts of the olfactometer that are more possible to get contaminated by residual smells. One of the most important properties when displaying smell is the ability to have short *onset* delivery times. Onset time is defined as the time needed from sending the digital signal to the olfactometer software until the concentration of the delivered odour reaches sufficient detection levels as measured by a sensor. For the purpose of measuring concentration levels in ppm, photo-ionization detectors are used. These are discussed in the following section (see section 2.4.4). Onset times are important when presenting smells with visual and/or auditory stimuli as it is necessary the multiple stimuli to be delivered synchronously to the user. Another desirable property of an olfactometer is the ability to deliver odour concentrations with limited fluctuations. This is important when a constant concentration stimulus needs to be delivered over a period of time. As pointed out previously, concentration level is linearly connected to odour perception (see section 2.4.2) thus it is important to deliver steady concentration values when the objective is to keep user's perception as uninterrupted as possible. The time interval between the onset time and the constant concentration delivery is usually termed the *rise time* of the olfactometer. It is required that this time interval is as small as possible so as the user to receive the correct concentration levels without time delays [LGA$^+$10]. It is also desirable the olfactory stimulus to be presented without any other tactile or auditory cues during the delivery [LEZP99].

Olfactory displays can be distinguished into two main categories. The first category includes olfactometers that make use of mass flow controllers for adjusting the injection of smell in liquid or gas form into a main air stream [Kob85, BP80, Etc80]. The injected smell is diluted into air producing an aromatised stream that is delivered to the user. In case that odour is not needed at some point, this type of olfactometer switches to delivering humidified air through the use of a control

channel while the pathways that inject smell remain close. This type of olfactometer has very good onset and rising times when presenting odour stimuli. This design requires the presence of a mechanism that can absorb residual smells and evacuate the tubes before presenting a new odour. This mechanism usually works by creating a pressure gradient at the tubes that need to be evacuated while the generated air stream is directed out of the olfactometer using an exhaust tube.

The second category has simpler designs where the odours are kept in small reservoirs without the need to be directed to a main air channel. These reservoirs normally comprise solenoid valves that can be controlled digitally. When air is directed towards one of the reservoirs, the valves automatically let it pass through the reservoir where is gets enriched with the odour that is stored there. After that the aromatised stream is directed to a container where it is connected with a control channel. At this stage the odour channel is sufficiently diluted before the delivery stage. The first olfactometer of this type was proposed by Lorig *et al.* [LEZP99] as a customised olfactory display that can be used to deliver smell impulses when conducting fMRI[2] scans in neuroscience experimental studies. Other examples of this type were implemented by Lundström *et al.* [LGA+10] and Lowen and Lukas [LL06]. This type of olfactometer is compact because the different components can be configured to be close or away from the user or from each other. This design is more advantageous for VR applications as it allows the coexistence of the olfactometer with other devices such as head mounted displays, audio headphones, tactile force feedback devices, etc. Other advantages of this design are the ease to maintain, reduced cost and the minimal effort that is required for its assembly. The disadvantage of this design is that it is more difficult to accurately estimate the onset and rise times [LGA+10].

There are many companies that manufacture computer controlled olfactometers that include many of the features described previously. The fact that there is no standard way to classify odours is a limiting factor towards constructing an olfactory display system that can be regarded as a widely accepted technological standard.

2.4.4 Methods for detecting and analysing smells

Photo-Ionization Detectors

Photo-Ionisation Detectors (PIDs) are sensitive sensors that are used to measure the concentration of an odour in an environment over time at any location of interest [Hon13]. This location is normally chosen to be the point that the odour enters

[2]Functional magnetic resonance imaging.

the nose. Gas molecules have a specific ionization energy level that is responsible for charging every atom of the odour and become an *ion*[3]. This amount of energy is termed ionization potential and is measured in electron Volts (eV) where 1 eV is approximately equal to 1.602×10^{-19} Joules. Different gases have different ionization potentials while tabulated data exist that present ionization potentials for every VOC [Bia16].

PIDs are able to detect the amount of atoms being charged due to the ionization potential for a range of different VOCs while remain insensitive for elements that are in gas form by nature and have large ionization potentials (e.g. oxygen). As mentioned earlier, these gases operate as "carriers" of the smell molecules. When odours are exposed to ultraviolet light their atoms are charged and trigger the PID sensor. In turn, the sensor, samples the aromatised stream and records amounts of charged atoms which give an estimation of the concentration at the sampling position. Commercial PIDs are able to detect VOCs in air at ionisation potentials below 10.6 eV whilst the maximum recorded concentration value ranges between 100 to 600 ppm depending on the PID sensor [Hon13].

PIDs sensors are useful for calibrating olfactory displays. Onset and rise times can be measured for different flow rates and using different VOCs so as to obtain an overview of what concentration values can be presented by the olfactometer and how these values fluctuate over time and under different parametrisations (e.g. flow rate, valve switching time, lower valve pressure, etc.).

Gas chromatography–Mass spectometry

As already stated, an odorous composition is usually comprised of two or more odorous components. Using chemical analysis of the odour is possible to detect these chemical substances and their proportions in the composition. This process is usually done using Gas Chromatography (GC) and Mass Spectometry (MS) analyses. The odour composition, normally in liquid form, is vaporised in a heated chamber and the produced gas is directed to a column inside the chromatography device where the different components are separated. The specification of the column depends on the expected features of the odorous molecules injected in the device, the temperature and the capacity of the column. This is a procedure that sometimes needs to be repeated multiple times and needs expertise so as to be carried out correctly. The separated compounds are directed into the mass spectometer where they are charged into ions using an appropriately modulated electric field. The ions are

[3]An ion is the state of an atom that the number of protons is different from the number of electrons.

subsequently detected and classified according to their mass to charge ratio values. The molecular mass and the mass to charge ratios are used to find out the identity of the original molecule, a procedure that is repeated for each of the compounds entered the mass spectometer. Other methods for analysing odorous solutions include coupled plasma and atomic emission spectroscopy techniques, or optical emission spectroscopy – mass spectometry.

GC and MS procedures for analysing chemical substances are considered the best possible technique for identifying the compounds of a substance and are widely used in odour industry. Detection of the chemical components that compose an odorous mixture can be used to replicate the smell or improve its hedonic features by introducing new components or simply by increasing/reducing the quantities of existing compounds in the mixture.

2.5 General rendering pipeline

This section introduces the basic pipeline used for simulating light, sound and smell transport in virtual environments. According to Greenberg *et al.* [GTS+99], rendering sensory stimuli includes the application of three basic stages. These are:

1. **Scene definition:** This stage includes the design of the scene CAD geometry, collect material and lighting information, capture anechoic sounds, define sound source and receiver locations in the VE, create mesh for smell transport simulation and define physical and numerical boundary conditions for the domain of interest. This is an essential step before proceeding to the next stages of the rendering pipeline.

2. **Transport simulation:** This stage involves the simulation of light/sound waves or the transport of the air-odour mixture in the VE defined from the first stage.

3. **Stimulus delivery:** At the last stage of the pipeline the visual (radiance), aural (acoustic pressure) and olfactory (concentration of the odour) estimated values are directed to the respective devices for delivery to the user.

2.6 Summary

The theory of this chapter mostly concentrated on the physical characteristics of light, audio and smell as well as the techniques for displaying these stimulations in VR frameworks. Light and audio are transported in the form of waves that

interact with various surfaces of the environment resulting to phenomena such as reflection, refraction, diffraction and interference. In the case of olfactory, diffusion and convection are the main physical properties that govern odour transport in a real environment thus simulating their effect is also desirable in VEs. The following chapter focuses on how stimuli are perceived by humans and establishes the link between the physical and the psychophysical interpretation for the senses of sight, auditory and olfactory. The material presented in this chapter is required for understanding of the rest of this thesis.

Chapter 3

Cross-modal perception and applications to rendering

The previous chapter introduced the reader to the main physical features of the senses of vision, hearing and olfactory. This chapter considers how human perceptual characteristics can be used to alleviate the large computational requirements needed for computing sensory cues of the three aforementioned senses. Although the human visual and auditory systems are very sophisticated, they have certain limitations that have been exploited to compute and deliver sensory stimuli in a non uniform way, i.e. compute a stimulus at a lower quality level without the quality degradation to be consciously perceived by a user.

In real life, multiple stimulations are received and processed simultaneously. Research in human perception has proven that the existence of multiple senses can result in perceptual interactions that can affect the way some or all of the sensory cues are perceived. These interactions have also been exploited to reduce the computation times in multi-sensory VEs. Interactions between the senses in VEs constitute an area of study known as *Cross-modal selective rendering*.

This chapter has a twofold purpose. Firstly, it describes how human selective attention operates both in a visual and an auditory context. Attention is the theoretical basis for computational saliency models which are used to identify parts of an image or audio track that are more possible to be attended by a human viewer/listener. Saliency maps have been extensively used to decrease computational times in graphics or auditory rendering applications. Secondly, the chapter reviews bi-modal interactions and the way they can be used to render one of the two modalities in lower quality without the user's perceptual experience being altered. As these perceptual interactions have typically only been examined in the context

of two senses (sight, hearing) this thesis moves knowledge forward and examines the effect of perceptual interactions in tri-modal VEs.

3.1 Anatomy of the human sensory system

This section presents the basic anatomy of the human visual, auditory and olfactory systems concentrating mainly on the organs that are used for the sensation of a stimulus. Stimulations are received by the eyes, ears and the nose and translated to signals that are transferred through the nervous system to the brain for further processing.

3.1.1 The human visual system

The human eye is composed of three tissue layers. These are the *sclera*, the *choroid* and the *retina* (see Figure 3.1). The sclera is the exterior white cover of the eye and it has an opening in front which is known as the cornea of the eye. The choroid is the intermediate layer between the sclera and the retina and is responsible for the formation of the iris in front of the eyeball. The iris is the colourful circular area in the eye that adjusts the diameter and size of the concentric circular area in the middle, known as the pupil. The pupil is responsible for controlling the amount of light that is allowed to pass into the interior of the eye and stimulate the retina that is located at the back of the cornea.

The retina is the curved surface that covers eye's surface interior. This surface is composed of ganglion and photoreceptor cells. Ganglion cells receive visual signals from the photoreceptor cells and transfer them to the head thalamus and hypothalamus for further processing. Two kind of photoreceptor cells co-exist in the retina. These are the *rods* and the *cones*. Rods operate at dark ambient conditions (scotopic region) and their spatial acuity is very low. In their operating range, rods are capable of detecting small variations of light intensity. Different from rods, cones are able to distinguish colour information and have much higher spatial acuity. There are three different types of cones. These mediate perception of long wavelength light (red), middle wavelength light (green) and short wavelength light (blue). Cones mainly operate at brighter light conditions than the rods, beginning from the photopic region [Fer01].

The foveal region or *fovea* is located in the center of the retina. This region contains the highest concentration of cones in the eye thus it provides the best possible visual acuity known also as *foveal vision*. The retina contains around 100

million rods located mostly in the periphery of the retina and around 5 million cones [Pug88].

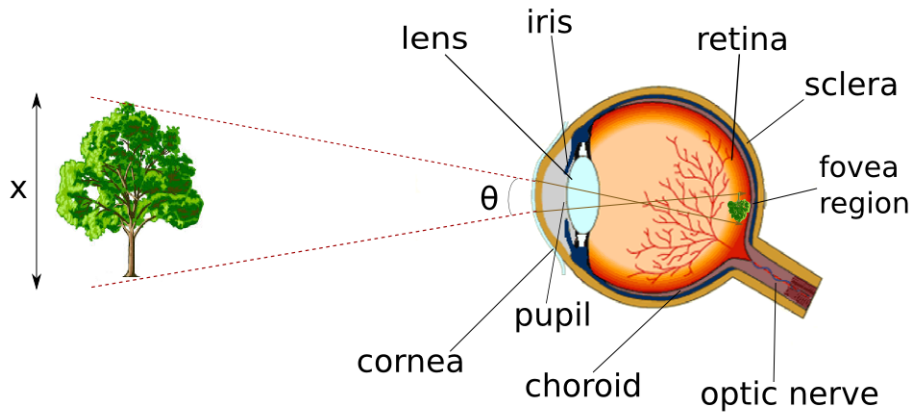


Figure 3.1: Basic anatomy of the Human Eye. Definitions and the original image can be found in Ferwerda's study [Fer01].

Although the human eye is a highly sophisticated sensory organ, it has specific limitations in terms of its visual acuity capability. These limitations refer to situations where the HVS cannot perceive very high resolution details or the temporal nature of the visual stimulus does not permit its full processing and perception.

In the spatial domain, when watching an object, one degree of the viewing scene is projected, across 288μm of the retina's surface where approximately 120 cones are located. This means that a repetitive pattern of alternating black and white stripes displayed at less than a degree of the viewing space cannot be perceived as a stripes texture anymore but rather as a blurred uniform grey region [Nag80]. This typical example shows that the spatial resolution of the HVS is limited. Depending on the distance d of the object away from the eye, the spatial resolution θ in degrees can be given as:

$$\tan(\frac{\theta}{2}) = \frac{x}{2d}, \tag{3.1}$$

where x is the height of the object in meters (see also Figure 3.1).

The temporal sensitivity of the human eye refers to the amount of information that can be captured and perceived when watching moving objects or imagery frames over time. In this case, the visual acuity is limited due to the inability to process the entirety of a dynamically varying scene. Information loss due to temporal restrictions can be partially mediated by the ability of the eyes to move (*saccadic*

eye movements) and foveate objects at high speeds in the visual domain. According to Henderson and Pierce [HP08], the human eye is capable of doing saccadic moves three times per second when watching a scene while the gaze fixation time at different scene locations can be at least 20 ms. During the fixation time, the eye foveates the object and extracts all its visual information (colour, shape, size, etc.). The temporal sensitivity of the human eye is 26Hz or 26 frames per second when watching a video [FN96]. The impression of continuous visual stimuli is attributed to a phenomenon known as *flicker fusion*. According to this, the retina persists perceiving the same image for a time interval of $\frac{1}{20}$ to $\frac{1}{5}$ after its first presentation [Rog25].

The number of photoreceptor cells in the retina is numerous compared to the ganglion cells that transfer signals to the brain. The HVS resolves this imbalance by selectively restricting the amount of visual information that is sent for further processing at any given time. Also the distribution of cones, which are more densely packed in the fovea region, enhances the selective processing of the visual stimuli [Fer01]. When an object is foveated by the eye, vision is sharper and more details are sent through the nervous system for further processing. On the other hand, peripheral vision, is achieved using photoreceptor cells that are not located in the fovea and provides blurred visual stimulations and less information is sent for further processing.

3.1.2 The human auditory system

The human ear is the sensory organ used to receive sound stimulations. Incoming sound waves cause vibrations that are translated to neural stimulations and transmitted through the auditory nerve to the brain for perception. The human ear is composed of three main parts. These are the *outer ear*, *middle ear* and the *inner ear* which individually serve different purposes in the process of capturing and processing auditory stimuli.

The outer ear includes the part that is known as the *pinna* of the ear (see Figure 3.2). The pinna contributes to sound localisation and acts as a reflector where the incoming sound waves are directed to the auditory channel. The pinna works as a band pass filter where only sounds within the audible range are allowed to pass to the middle ear. The selective behaviour of the pinna to discard parts of the frequency spectrum is termed *pinna notch* behaviour [SAD07].

The middle ear is the connecting region between the outer ear and the inner ear. At this part, the ear drum is located at the end of the auditory canal. As the incoming sound triggers the ear drum, it make it oscillate and, in turn, these oscillations are transmitted through the ossicles to the oval window. The ossicles

are small bones connected in such a way so as to amplify the received vibrations before reaching the oval window which is a patch of tissue that is used to transmit the vibrations to the inner ear. The middle ear operates at the same atmospheric pressure that is present at the outer ear. If for any reason a pressure gradient exists between the outer and middle regions of the ear, a pressure equalisation operation takes place through the use of the Eustachian tube that connects the throat with the middle ear [MC15].

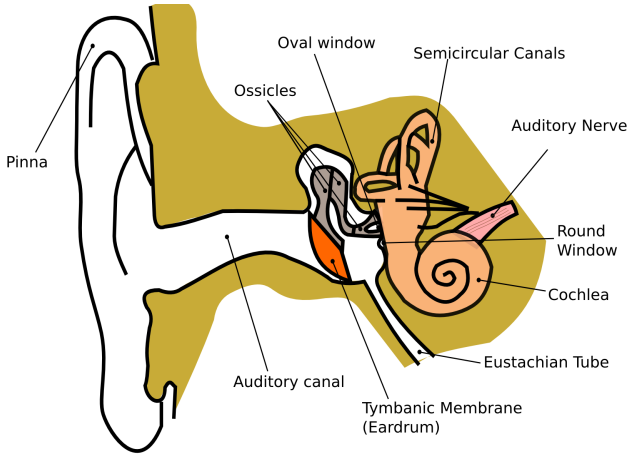


Figure 3.2: Basic anatomy of the Human Ear. Definitions and the original figure can be found in Frederick *et al.* study [MC15].

Lastly, the inner ear is a liquid filled region that contains the core of the human auditory system, the *cochlea*. The basic component of the cochlea is the Organ of Corti which is a coiled shaped tube that contains the basilar membrane. This membrane is about 3.5 cm long and contains fibres from the base all along its apex. The basilar membrane operates selectively at different incoming frequencies and particularly it vibrates at different regions. For high frequencies ($5 - 20$ kHz) the base part vibrates which is located at the beginning of the cochlea while at middle and lower frequencies the centre and the apex are activated respectively. The role of the basilar membrane is the decomposition of the incoming sound to fundamental frequencies, an operation that can be intuitively thought similar to a Fast Fourier Transform (FFT) [Ste11]. The cochlea is the stage where the sound wave is transformed to neural signals and transmitted to the brain through the auditory nerve.

The human auditory system does not possess the same spatial acuity as the human visual system. Specifically, Perrott and Saberi [PS90], implemented an experiment to find the smallest azimuth and elevation angles that a human listener

perceives that a sound stimulus is coming from a different direction. Their results showed that the smallest azimuth angle is approximately 0.97° while keeping the elevation angle constant. The smallest elevation angle without varying the azimuth was 3.65°. As far as the temporal sensitivity of the HAS is concerned, this is 89.3Hz [FN96] and is significantly higher that HVS's temporal acuity (26Hz). The temporal sensitivity of the HAS is responsible for perceiving an auditory illusion is known as the *continuity illusion* [WWP88]. Specifically, if a sound is turned off for a while and replaced by another sound, the first sound is perceived as continuous for a short time period. This is analogous to the flicker fusion of the retina in the visual domain.

3.1.3 The human olfactory system

Of all the senses, smell is the least understood. Everyday we sense a vast number of smells that provide valuable information regarding the surrounding environment. Scents can influence people's reactions while there is evidence that people strongly recall memories associated with a perceived scent, a phenomenon known as *Olfactory-evoked memory* [HE96]

The human olfactory system is composed of the main olfactory system and the accessory system. The main olfactory system is responsible for detecting volatile compounds in the ambient air whilst the accessory system is mostly connected to the sense of taste as it is capable of detecting smells in liquid form when we taste. It should also be noted that olfaction and taste (gustatory system) are closely connected senses.

In the main olfactory system, smells are inhaled together with air through the nostrils in the form of a gas mixture. This mixture approaches the *Olfactory epithelium* where microscopic hair fibres, known as *cillia*, are used to capture smell molecules from the mixture and dissolve them into the mucus of the epithelium (see Figure 3.3). The olfactory epithelium is a layer of tissue of around $4 - 10$ cm^2 located in the nasal cavity, approximately 7 cm to the back of the nose. After the dissolving, compounds diffuse upwards where they are detected by a range of membrane proteins which are located at the receptor cells. According to Bushdid *et al.* [BMVK14], there are over 40 different million receptor cells in the human epithelium.

Receptor proteins are located at the receptor cells and bind to the incoming smell molecules. This binding is specific to the active site of each protein type and the receptor neurons convert the molecule–protein binding into electrical signals using a process that is not fully understood by the research community [AY05]. After

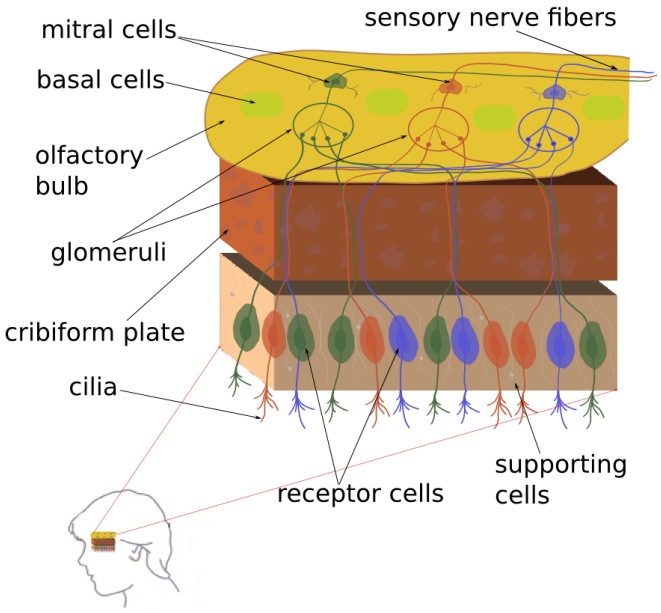


Figure 3.3: Basic anatomy of the human olfactory system. Definitions and the original figure can be found in Purves's *et al.* study [PAF01].

this phase, the electrical signals pass through the cribiform layer, a thin bone layer, and travel trough a system of local neurons to collection locations (or glomeruli) located at the olfactory bulb. Apart from glomeruli structures the olfactory bulb also contains cells, known as *Basal cells*, that are used to periodically create new receptor neurons every 2 to 4 weeks. At that point, the neural signals are forwarded to the mitral cells which are the starting points of the nerve olfactory neurons. These neurons transfer the signals to the Cerebral cortex, an area of the brain that is used for the recognition and perception of the incoming stimulus.

An important feature of olfaction that distinguishes it from the other senses is that sensory signals are transferred to the cerebral cortex directly without first passing through intermediate stages (thalamus and hypothalamus) as is the case with all the other senses [SMKK01]. Many studies associate the transfer of neural signals through the thalamus with the existence of selective attention [DS98], a theoretical topic from Psychology that is discussed in the following section.

3.2 Selective attention

Selective attention is a psychology term which is used to describe the process of selecting relevant sensory information for further processing while ignoring all the irrelevant information [Bro58, LFR04]. For example, in the visual domain, although

the eye receives a plethora of visual information, only a portion of it is actually attended and fully processed by the brain. Attention is not a perceptual phenomenon restricted only to vision [Yar67] but exists also to other modalities including auditory [ASA12], olfactory [SMKK01] and somatosensory [LBHM99] domains. Many different classifications for perceptual attention have been proposed in the past. The interested reader can go through the study of Chun *et al.* [CCTB11] where a wide classification scheme is proposed. This scheme encompasses spatial, temporal, object based and feature based perceptual attention. Also the authors consider attention due to a modality when multiple senses are considered at the same time. Attention is the cornerstone of the visual and auditory saliency computational models while it constitutes the theoretical background of various perceptual metrics that are frequently used in computer graphics.

3.2.1 Visual attention

A broad distinction of attention in the visual domain can be made between the *Bottom-up* ([Jam80]) (also known as *exogenous* or *passive*) attention and the *Top-down* [DD95] (also known as *endogenous* or *active*) voluntary shifting of attention. The former form of attention refers to the involuntary gazing to potentially conspicuous stimuli in the environment, for example a car crash in front of us can definitely attract attention. The top-down attention refers to the voluntary focus of attention towards important visual stimulations not because they are salient but because they are considered as a"goal". For example, the top down mechanism allows to read a magazine while sitting in a crowdy place. Top-down and bottom up are complementary perceptual mechanisms and they do not work in an exclusive fashion. If attention was working solely in a bottom-up way then humans would never develop a behaviour directed by aims as salient stimuli would constantly distract them. On the other hand, even when focusing on a target, our attention still requires to be attracted by salient features of the surrounding environment.

Many researchers have reviewed the different characteristics of the bottom-up and top-down forms of perceptual attention [IK00]. On the one hand, bottom-up attention cannot be restricted, it is not affected by person's expectancies and previous memories experiences and does not result to conscious awareness. The top-down attention model, on the other hand, is resource limited, it can be easily manipulated and it is strongly affected by a person's previous memories and expectancies. Top-down attention results in conscious awareness of visual cues.

Yarbus, in 1967 [Yar67], conducted a series of experimental studies to investigate the effect of top-down perceptual attention compared to the bottom-up

approach. Specifically, he used an eye tracking device to record participants' eye trajectories when viewing an image (a copy of Repin's painting titled "The unexpected visitor") under different experimental conditions. Subjects were asked to view a picture while they were assigned a specific task related to the content of the picture (e.g. to determine the ages of the characters that are depicted on the image or remember their clothes). Apart from the experimental trials that included a task, there were trials where participants were simply asked to freely examine the painting. The eye tracking results demonstrated that participants fixate on relevant objects when assigned a task compared to the case of freely viewing the image. This experiment clearly showed the distinction between top-down and bottom-up models of attention.

Visual attention can also be discriminated into *Overt* and *Covert* shifts of attention [Pos80]. Overt shifts of attention describe the redirection of the eyes towards areas of interest or salient objects. The human eye usually performs overt shifts using saccadic eye movements and fixation time intervals. On the other hand, covert attention does not involve eye movements and it enables faster response to attended visual events. Overt attention differs from covert attention in the same way that gazing at an object is different than observing something with our peripheral vision (the corner of the eye). Experimental studies have measured reaction times (RT) in both covert and overt shifts of attention in either uni-modal scenarios or cases where multiple senses are delivered at the same time (multi-modal scenarios) [WW08].

Another psychological phenomenon that frequently occurs in the visual domain is a person failing to notice an unexpected stimulus that is in plain eyesight [RLGM92]. This phenomenon is known as *Inattentional Blindness* and it is not caused by physical inadequacies of the human eye but rather it characterises the failure of perceptual attention when watching multiple visual stimuli. This phenomenon was discovered and named by Mack and Rock in 1998 who claimed that the stage of consciously perceiving a stimulus requires first to attend it [MR98]. Specifically, the authors argue that the perception of a shape when watching a visual stimulus requires visual attention while the perception of colour, location and motion features do not necessarily require attention to be consciously perceived. In another study, by Simmons *et al.* [SC99], the authors demonstrate a practical example of Inattentional Blindness through a task where participants watch two basketball teams (black and white) in a video to change ball passes. The task is to count how many passes the white team made. During this task, participants miss to attend to a gorilla or a woman with an umbrella passing through the scene. In the first demonstration of the experiment only 8% of the participants were able to

report the presence of the unexpected character in the video.

An equally significant phenomenon that is related to human perceptual attention is known as *Changed Blindness*. In this case, the HVS fails to notice visual differences that are included in a visual scene or a digital image. Typically, for changed blindness to occur, the change in the scene needs to coincide with a visual distraction, for instance, a focus of attention to a different spatial position [ROC97]. The HVS does not capture the newly included information and the differences remain undetected by the viewer.

3.2.2 Auditory attention

Research in the 1950s concentrated primarily on selective attention in the auditory domain and specifically the interpretation of the so called *Cocktail party* problem. When many sounds enter the ears at once the HAS is able to separate out the sound of interest from the mix of auditory stimuli either because it contains salient features or because it is something that we expect to hear as part of a previously assigned task or objective. One of the first researchers in selective auditory attention was Cherry in 1953 [Che53]. He explored the cocktail party problem using an experimental method termed as the "selective shadowing task". In its original form, two different auditory stimuli are delivered through headphones and listeners were asked to repeat out loud one of them ("shadowed" message). Cherry noticed that participants were unable to report the contents of the non-shadowed message after the end of the experimental session with two important exceptions. The first exception was the case of hearing their own name as the non-shadow stimulus and the second, included the hearing of sounds of significantly different pitch than the ones received in non-shadowed message. In the latter case, the pitch differences immediately attracted listener's auditory attention.

Another important aspect of the auditory attention is the ability to focus on salient auditory stimuli in a selective fashion. This was investigated by Alain and Arnott [AA00] who noticed that the HAS is able to receive a cluttered auditory input and keep only the portion that is considered meaningful auditory information. This feature of the HAS is known as *Auditory scene analysis*. The auditory system captures a cluttered auditory stimulus, tries to separate it to different sound streams according to specific psychological principles [Bre94] (these are known as Gestalt's principles) and then passes it to higher levels of processing. At this point, auditory attention selectively captures some of these streams and processes them through elaborate methods (known as *object-based hypothesis*) that lead to the cognition of the incoming auditory stimulations.

3.2.3 Olfactory attention

Olfactory attention is a topic that has not received much merit compared to perceptual attention at other sensory modalities. Many researchers claimed that the direct transfer of the neural signals to the cortex without passing first through the thalamus of the brain (see section 3.1.3) might suggest that humans are not able to selectively attend olfactory stimuli [MT94, Smy97]. In an experimental study by Charles *et al.* [SMKK01] the authors prove that olfactory stimulations can be selectively attended. Their experimental design included the presentation of both visual and olfactory stimuli and participants had to judge if the presented stimulus was "weak" or "strong" in terms of intensity by pressing one of the two supplied pedals. Before the stimulus presentation, an auditory cue of different pitch primed the subjects if a visual or an olfactory stimulus would follow and in most, but not always, of the cases the subjects were primed correctly. The results showed that subjects responded faster when they were primed with the correct auditory cue. This indicated that they were attending to the arrival of an olfactory stimulus.

A more recent study by Keller [Kel11] examines whether olfactory attention can be classified in a general attention taxonomy as is the case with visual and auditory attention (see [CCTB11]). Specifically, the author argues that olfactory attention possesses spatial characteristics in the form of overt shifts (turn the head to sense different odours) while temporal olfactory attention happens but with interruptions as inhaling is not a continuous process but periodical.

Another form of olfactory attention that is discussed by Keller [Kel11] is attention to olfactory objects and their features. Keller notes that there is some ambiguity of what is defined as an "olfactory object" with possible interpretations to be an odour source or an odour that is released from a source or even how pleasant or not an odour is. For this reason, he does not accept that olfactory attention can be shifted to different olfactory objects but rather it is attributed to one object at a time. In vision, for example, attention might be shifted towards different objects in the visual domain depending on their saliency and the observer's task. Another form of attention is feature based attention and can be defined using two different ways. The first way is attention to specific features of an object (e.g. colour, shape, orientation in vision) and the second is attention to a specific quality of one feature. In olfaction, feature based attention can be of the latter form. For example, when smelling a glass of brandy, in most cases, people are able to attend whether the perceived odour has a woody or fruity smell.

Finally, Keller examines if attention can be shifted between modalities including shifts due to olfactory attention. In the visual and auditory modalities attention

can be shifted and this has been shown in an earlier study by Spence and Driver in 1997 [SD97]. It turns out that attention can also be shifted due to the presence of an olfactory stimulus. The latter point was demonstrated in an experimental study by Spence *et al.* [SMKK01].

3.2.4 Discussion on attention

An important question that frequently arises is whether failure to attend to a stimulus can be directly attributed to the underlying perceptual mechanism or other external factors. Specifically, it can be claimed that lack of attending to a stimulus is due to people's limitation to remember events that happened in a short time interval. An alternative explanation can be based on the fact that people might interpret the incoming stimulus differently and this interpretation does not match the strict criteria of an experimental procedure that evaluates attentional capture.

In a study published by Wolfe and Horowitz [Wol99] many of these considerations have been reported and discussed in the context of Inattentional Blindness (see section 3.2). The event that someone can attend to a stimulus and forget it until the time of the experimental evaluation (completing a questionnaire or interviewed) is quite possible and does not necessarily imply that he/she did not attend to the delivered stimulus but simply could not recall it after the end of the experimental session. Furthermore, in the case of examining inattentional blindness using the selective looking example, it is quite possible that people interpret the presence of the unexpected visual object (e.g. gorilla in [SC99]) in a slightly different form other than the one the experimenter would expect. This is known as *Inattentional agnosia*. For example, someone might have perceived the gorilla as a shadow or a player of the black team that ducks for a while on the floor. Also it should be said that a participant who has no clear interpretation of the event might be hesitant to answer whether the stimulus was attended or not as he/she might be unsure of its exact content.

The experimental method that is used to evaluate attention is also another point that merits further discussion. The fact that the human attentional mechanism is so complicated, raises questions on whether or not the experimental evaluation (e.g. verbal description of the attended stimulus) is able to infer if all the presented information was attended to or only a portion of it. All the experimental methods use some sort of performance measure to evaluate if the presented stimulus was actually attended or not. This performance metric might underestimate the amount of information that has been attended. Even in the hypothetical case that the performance measure is perfectly tailored for the experimental procedure,

it should not be forgotten that some other underlying perceptual mechanism, and not attention, might be contributing to attend a portion of the incoming sensory information or even mask the information that is meant to be attributed to the attentional mechanism.

Another question that remains open in the literature is whether the perceptual attention mechanism works in the same way when dealing with more complex stimuli. For example, stimulations inspired from every day events and occurring in the surrounding environment would probably attract attention more both in an implicit or explicit fashion. All the studies presented so far considered measurements of attention within strict laboratory conditions but there is no guarantee that the same arguments hold when the incoming stimuli have complex attributes in one or multiple modalities at the same time. Another point that still remains unexplored is how attention is allocated when more than one salient object is present at the same time or when more complex tasks (possibly across different modalities) are assigned to the human subjects. The criteria that determine attention in these cases remain vague.

3.3 Cross-modality in human perception

In every day life, people sense a multitude of different sensory stimuli that originate either from a common or multiple sources and affect all or some of the individual sensory systems. Typically, the senses are designed to operate in a conjoint fashion when receiving and processing multi-sensory input. This leads to an overall understanding of the surrounding environment. For instance, at a crowded place, auditory selective attention is used to perceive what other speakers say (cocktail party phenomenon) but visual cues such as lips movement, face or body gestures are also gathered and account for enhaning the perception of the incoming auditory stimulus.

Many psychological studies have proved that stimuli reaching our sensory organs are not processed independently [TN03b, BR81], instead, humans incorporate multi-modal information in order to produce a full cognition of their external environment [Duf99, SD04, CSS04]. Even in the rare case that one modality cannot provide useful sensory information, the others assist in order to receive information for a task completion or for the comprehension of an event.

As has already been mentioned, the individual sensory systems have different spatial and temporal sensitivities. These discrepancies are the main reason for the perception of illusionary effects. A typical example is the *ventriloquism effect* [HT66]

in which the viewer is fooled into thinking that sound emanates from a visual cue.

To be able to understand cross-modal interactions, it is important to take into account the *modality appropriateness hypothesis* [WW80, BKT86]. This hypothesis advocates that depending on the nature of a task, the most appropriate modality to the task's context will prevail the process of perceiving a multisensory stimulation. Vision has higher spatial resolution, thus it prevails in spatial tasks whilst audition has higher temporal resolution and as a result it prevails in temporal oriented tasks. Furthermore, the sense of smell can influence our behavioural performance and cognitive operation, therefore it is important to see how olfactory stimulations are perceived along with other senses. Psychological evidence claims that olfactory stimuli are strongly involved in cross-modal information processing [SRMN10, MBM02]. This fact makes the potential links of olfaction with other senses very substantial.

This section refers to cross-modal interactions between vision and audition as well as between olfaction and vision. In each of these pairs, the way one modality can be affected under the impact of the other is also taken into account yielding bi-directional interactions.

3.3.1 Effect of hearing on sight perception

Even though sight is considered a dominant modality and there are many known phenomena that explain how it affects other modalities, there are examples which show how perception of visual stimuli can be affected by the presence of sound distractors.

A study by Scheier *et al.* [SNS99] demonstrated how sound alters the temporal features of a visual stimulus. Their experimental framework included two LED lights that were at slightly different locations and turned on with a time delay of 60 ms. They found that subjects were able to make better judgements of the sequence that the lights were turned on when a sound was preceding the presentation of the first light and a second sound was following the second light (i.e. A-V-V-A). On the other hand, when two sounds were heard between the visual stimuli (i.e. V-A-A-V temporal order) participants' performance was decreased significantly. In that case the temporal sensitivity of the audio affects the perception of a visual stimulus.

Shams *et al.* [SKS00] demonstrated the *illusory flash effect*, which shows how multiple sounds can affect the perception of a visual cue. The experiment included the presentation of a spot on a screen flashed once and accompanied with one to three beeps. When a single flash was accompanied by multiple auditory beeps, it was perceived as multiple flashes (see also Figure 3.4). The authors explored the

possibility that the illusory phenomenon was induced due to the difficulty of the task or some cognitive bias caused by the sounds. Using catch trials and control conditions they verified that this was indeed an illusion. This study also revealed that modalities and events are asymmetric to each other. This means that visual perception is affected by sound only when a single flash preceded multiple beeps, but in the case a single beep preceded multiple flashes auditory perception remained unaffected.

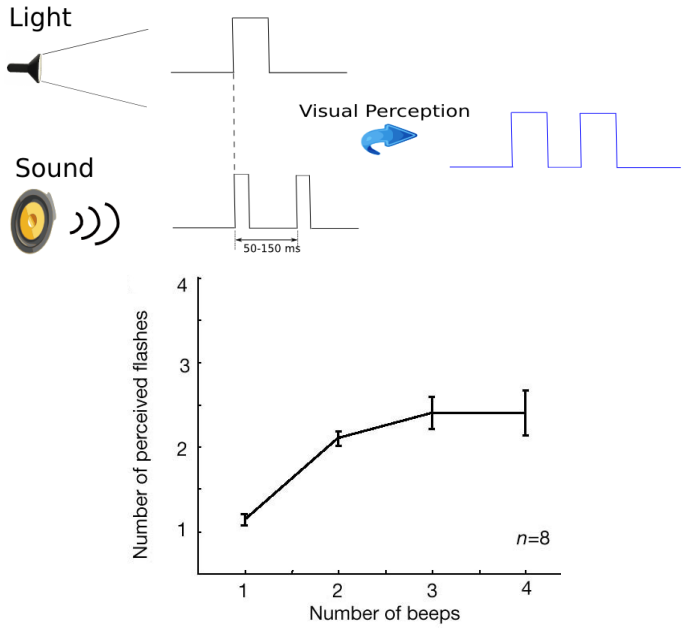


Figure 3.4: **Top:** Graphical representation of the Illusory flash effect. Synchronous presentation of a visual and two auditory stimuli creates the perceived illusion of two visual stimuli. **Bottom:** Experimental results, adopted from [SKS00], show how the number of perceived flashes scales when multiple beeps follow the presentation of a single visual stimulus.

Another interesting study explored how the motion of an auditory stimulus can possible affect the perception of a moving visual stimulus. In this study [MW01], a visual dot conducts a random move across a display whilst a binaural auditory tone is delivered to the user using headphones. The auditory stimulus carries out a horizontal motion from the left to the right ear and vice versa. The experimental results indicated that the auditory motion strongly biased the perceived trajectory of the visual stimulus.

3.3.2 Effect of sight on hearing perception

The effect of visual stimulation upon the perception of auditory cues has also been extensively studied in the context of cross-modal perception. Phenomena such as the *ventriloquism effect* or the *McGurk* effect constitute conditions which confirm that the perceived spatial location of a sound stimulus can be affected under the impact of a visual stimulation.

The ventriloquism effect is produced when a visual stimulus is presented concurrently with an auditory stimulus but at slightly different locations [HT66, CWGJ75]. The result is that both stimuli seem to originate at the location of the visual stimulus. It was termed, *ventriloquism effect* by Howard and Templeton [HT66] as it was reminiscent of the illusion created by ventriloquists when they produce speech without any lip movements while the words are perceived to come out from a puppet's mouth. A similar illusion is experienced when watching television. The visual and auditory outputs of the device are in different places but the viewer perceives the sound as coming out from actors' mouths. Many studies have explored the necessary conditions to replicate the perception of the ventriloquism illusion. These studies demonstrate that spatial and temporal disparities between the two modalities can enhance or fade the phenomenon when they are adjusted appropriately [TC73]. The ventriloquism effect is a particular example of a more general phenomenon known as *visual capture*. According to this, humans tend to use sight as the preceding modality for perceiving the world when multiple stimulations are present at the same time.

Visual capture does not refer only to the modulations of audio by sight but it is also used to describe the change of proprioception sensory stimulations by sight. Experimental studies investigated the percept of limb placement under conditions where visual perception is shifted at many different areas of the visual space (this is usually done using prism spectacles) [HPI14]. The results indicated that subjects tended to mislocalise their limb position when they received a displaced visual stimulation even when they were advised to use hand movements to make their judgements.

Another example that shows how visual cues can affect auditory perception is the *McGurk* effect that discovered by McGurk and McDonald in 1976 [MM76]. The McGurk effect is also a typical example of visual capture. In this phenomenon, a person says the word "ba" while the observers perceive the word "fa". This happens due to the lip movements. When someone closes his/her eyes the word "ba" can be heard clearly.

3.3.3 Effect of sight on odour perception

It is well known that humans do not only count on information coming from olfactory sensations in order to identify an odour. A study by Desor and Beauchamp [DB74] reports that people were able to identify one third of the odours presented to them without using stimuli coming from other modalities. The recognition of an odour can be significantly facilitated using visual information regarding the colour of the liquid mixture that emits the odour or even better when reading its package label. The colour of a liquid odorous solution can significantly prime its identification [SIS+05] because it narrows down the number of scents that someone remembers to a small subset of smells closely related to the perceived colour. For instance, the orange colour of a solution primes someone to think that the perceived smell comes from orange juice and not a differently coloured beverage.

Researchers have reported that odour perception can be significantly affected by relevant visual stimuli [GRK96, MBD01]. Sakai *et al.* [SIS+05] found that all subjects correctly identified a cola-flavoured beverage when it was coloured dark brown but many failed when it was coloured orange. Morrot *et al.* [MBD01] conducted a similar experiment using red and white wines. They found that red-coloured white wine was misidentified as being red wine by all 54 participants who participated in their experiments. It is worth noting that all the participants were oenology undergraduate students. In another experiment, Gottfried and Dolan [GD03] asked subjects to detect an odour while they were examining an image. The image and the smell stimuli could be either congruent or incongruent. Their results showed that watching congruent pictures facilitated the odour identification as compared to incongruent stimulations or trials where no visual stimulus was delivered.

Lastly, Degel and Köster [DK98] conducted an experiment where subjects were asked to match a set of odours kept in containers with a set of visual stimuli in the form of images. Participants successfully matche, on average, any odour stimulus with the image that is most closely related to it. This study shows that the presence of odours can bias subjects' choices in the visual domain.

3.3.4 Effect of odour on sight perception

Having considered the influence of visual stimulations to odour perception it is also reasonable to examine the impact of olfactory stimuli on visual perception. This perspective has not been investigated much. One of the first studies that explored this direction was by Knasko in 1995 [Kna95]. This study investigated the effect of odours when processing images. Specifically, participants were exposed to pleasant

ambient odours while watching a sequence of slides. Experimenters noticed that people devoted much more time watching a slide when exposed to a smell congruent to the slide as opposed to an incongruent scent stimulation or when no odour was present at all.

In a more recent study, Seo *et al.* [SRMN10] extended these results and examined whether or not an odour can selectively influence visual attention towards a congruent object. As opposed to previous studies, the authors measured participants' fixation times when watching different visual objects. Their findings clearly demonstrated that subjects tend to attend a visual stimulus more often and for longer time under the effect of a congruent olfactory stimulus compared to the case of an unrelated odour stimulation.

In another study, Seigneuric *et al.* [SDJ+10] implemented an experiment in which participants freely explored images that contained familiar visual cues while at the same time an odour congruent with one of these images was emitted without subjects to be aware of its presence. Experimenters made use of an eye tracking device to investigate how the eye fixations are distributed across the presented objects and for how long. They found that participants fixated less time on odour related objects compared to the smell unrelated objects. Also, the time needed until the first fixation on an odour related visual object was significantly reduced. These results suggested that a familiar odour can assist the identification of a congruent visual object. Furthermore, odour stimuli affected subjects to swiftly their gaze towards the odour related object confirming that olfaction can bias visual attention towards a congruent visual object.

3.4 Cognitive resources and limitations

The human attentional mechanism has specific limitations and cannot be fully functional at all times. In the visual domain, as it was discussed earlier (see section 3.2), Inattentional blindness is an example which shows that people cannot discern a salient object because their cognitive resources are involved in carrying out an unrelated task. In the same way, within the auditory domain, there is the cocktail party phenomenon where people segregate an interesting sound signal in a cluttered environment depending on its saliency or an assigned task. This implies that perceptual resources are employed for the successful implementation of the task and other related audio or visual stimulations are ignored.

According to Pashner [Pas99], human attention has limited capacity and regardless of whether a stimulus is attended voluntary or not it is a process that

may lead to fatigue. Although, it is well known that attention is limited, there is no study that explicitly investigates into how to quantifying the attentional limits of the human perception. On this topic two opposing opinions have been formed in the literature. The first supports that attentional resources are shared across all the senses with no exception [SRD00, SJ01]. This viewpoint is known as the *Inter-modal* approach. The second opinion supports that every sense uses a dedicated repository of attentional resources, known as the *Intra-modal* approach [AAR72, AMB06, DD06]. The process of assessing whether the first or the second viewpoint is true depends on many criteria including, but not limited to, the detection versus discrimination paradigm and the role of the short-term memory in human attention [MW77].

3.4.1 Inter-modal

This class of perceptual models claims that attention resources are common for all the senses and performance on a task can be affected when a stimulus in some other modality requires our attention. For instance, using a cellphone while driving a car can disorganise one of the two tasks because attention needs to be divided [SJ01] across the two processes. Also, there is evidence that attending to different spatial locations at two different senses can be very difficult. Driver *et al.* [DS94] demonstrated the validity of this hypothesis using an experimental study where audio was delivered to the user either from the left or the right side of the head. In addition to sound, lip movements were presented to the user on either the same or opposite sides of the incoming sound direction. These movements were either synchronous with the audio track or asynchronous with its content. The results showed that when sound was presented on the same side with synchronous lip movements the subject performed considerably better in attending both of the delivered stimuli.

3.4.2 Intra-modal

A contrary viewpoint regarding the human attentional resources has been supported by Alais *et al.* [AMB06]. The authors showed in their experimental study that humans were able to attend to low level visual and auditory stimuli without facing any performance issues. Therefore, they set again the question of whether attentional resources are actually shared between the sensory modalities or not. They showed in their results that people were able to perform equally well when they were assigned either a single stimulus or a bi-modal dual task. Nonetheless, when subjects were assigned tasks that required attention at the same modality, performance was significantly reduced implying that attentional limitations exist in the context of the

same modality.

3.4.3 Cross-modal integration

Burr and Allais [DD06] proposed a framework in which cross-modal information can be combined as a sum of all individual stimuli estimates weighted appropriately. The estimate can be calculated as $\hat{S} = w_A\hat{S}_A + w_V\hat{S}_V$, where w_A and w_V are weights by which the individual stimuli are scaled, and $\hat{\mathrm{S}}_A$ and $\hat{\mathrm{S}}_V$ are independent estimates for audition and vision respectively. The weights are inversely proportional to the auditory and visual variances (σ^2) of the underlying noise distribution $w_A = 1/\sigma_A^2, w_V = 1/\sigma_V^2$. Their proposed model has been tested using different visual stimuli with different levels of blurriness [AB04]. An example where audition captures the sight occurs when visual stimuli are corrupted by blurring the visual target over a large region. The blurring, however, has to be significant i.e. over about 60°, which makes most scenes unrecognisable. In this study, auditory localisation was performed using only inter-aural time difference which account only for one class of available sound localisation cues (see also section 2.2.3).

Chalmers *et al.* [CHM09, CF10] proposed an extension of this model to encompass multiple senses in VEs. The authors argue that the ideal virtual environment should provide an overall perceptual experience that is equivalent to a real world environment. This can be achieved when stimulations from multiple senses are combined together appropriately. Specifically, the overall perception denoted as $P(t, \rho)$, where t is the effect of the assigned task and ρ is the effect of preconditioning, can be written as a linear combination of all sensory stimuli as:

$$P(t, \rho) = \omega_V V + \omega_A A + \omega_S S + \omega_T T + \omega_H H + C, \tag{3.2}$$

where the weights $\omega_V, \omega_A, \omega_S, \omega_T, \omega_H$ define the necessary fractions of each of the five senses (Visual, Auditory, Olfactory, Gustatory and Haptics) for creating a perceptually equivalent analogue of a real experience. The authors explain that the selected weight values are the minimum required proportions for each sensory stimulus and higher values would not have any perceptual benefit but would unnecessarily increase the computational requirements. Although this theoretical approach summarises all the required components for cross-modal integration, further work is required to validate if the linear relationship of the independent variables offers more advantages compared to any other algebraic relationship.

3.5 Saliency

Saliency refers to areas or objects that attract attention in visual or auditory modalities. In the visual domain, salient objects are likely to attract humans' gaze towards them while in the auditory domain, saliency refers to sound features that acoustically "stand out" from an auditory input. Saliency can be quantified using saliency maps both in vision and acoustics. These are images that encode how salient a region is by attributing brighter pixel values at these regions compared to non-salient areas where darker pixel values are assigned.

Saliency models have many applications in surveillance, human or object recognition, artificial intelligence applications, Virtual Reality (VR), etc. They have been extensively used to reduce computation times in selective rendering frameworks where an image or an audio is rendered in a non-uniform way following the information encoded in a saliency map. This means that areas that include salient features are rendered at high quality whilst non salient regions are assigned moderate or lower quality. Any algorithm that needs to extract visual or auditory salient information from a complex input uses some form of saliency model.

3.5.1 Visual saliency

The notion of *visual saliency* was presented for the first time in a neurological model introduced by Koch and Ullmann in 1985 [KU85]. Their idea was to find all the visually conspicuous locations in a scene and extract them in descending order of saliency. The first implementation of a saliency model was proposed by Itti and Koch [IKN98, IK00] and replicated to a certain degree how the HVS extracts salient information when attending a visual scene. Their work resulted to the construction of a computational model that provides what is known as a *Saliency map*, an image that serves as a guide to predict areas that are most likely to be attended by a human observer when an image is given as input to the model. An example of an input image and its saliency map is given in Figure 3.5.

The saliency computational model presented by Itti and Koch includes the use of intensity, colour and orientation as primary features for detecting salient regions in an image. The model is composed of four stages. In the first stage the image is filtered using a series of Linear and Time Invariant (LTI) filters (see Appendix B) at a range of different resolutions starting from low frequency images (blurry images) up to the original image. This stage produces a series of intermediate images that can be grouped to four colour channels (Red, Green, Blue, Yellow), an intensity channel and four orientation channels (using angles at direc-

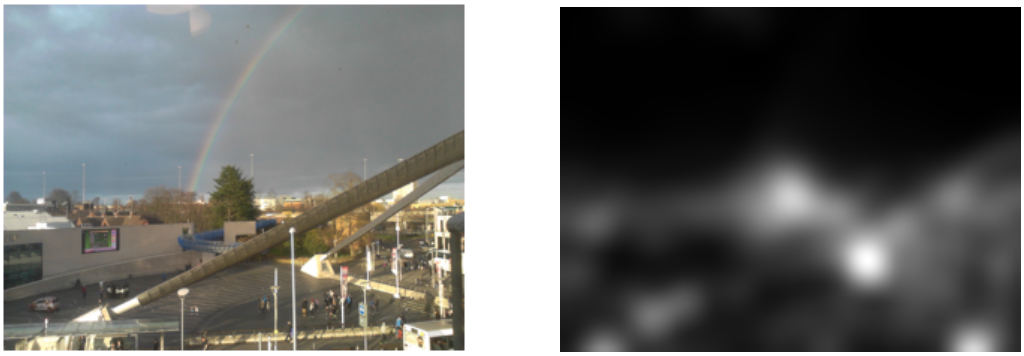


Figure 3.5: Original image and the computed visual saliency map. The saliency map was implemented following Itti and Koch's model [IKN98]. The original image was captured by the author.

tions: $0°, 45°, 90°, 135°$). The second stage computes differences between different resolution images to imitate the property of the eye receptors to distinguish a visual cue when it significantly differs from its surroundings. This stage is termed as *center-surround differences* and results in a range of images at different scales that encode exactly this behaviour. These images are termed *feature maps*. In the third stage, the feature maps are normalised so as to limit the noise and promote the areas of larger center-surround differences. This stage includes a linear combination of all the feature maps and yields the saliency map. The last stage includes the application of a repetitive technique to find and select the most salient locations in the saliency map. This is implemented using a "Winner Takes All" (WTA) neural network mechanism. The WTA selects the current most salient location and suppresses it for some time such that another salient location can be promoted and selected. The WTA methodology mimics a well known psychophysical mechanism of human perception known as *inhibition of return* [IK99]. According to this, the human eye identifies the most salient location and temporarily inhibits it in order to scan other possible conspicuous locations in the visual field.

Meur *et al.* [LMLCBT06] improved the original saliency model by Itti and Koch by implementing many other psychophysical properties of the HVS. As opposed to the original saliency model this model computes hierarchical perceptual representations of the visual input giving more accurate saliency estimations compared to the Itti and Koch computational model. In particular, this new model takes into account contrast sensitivity functions, perceptual decomposition, visual masking and center-surround interactions as extra processing modules for better saliency estimations. The authors investigated the accuracy of the saliency estimation model by using case studies where eye-tracking results were collected and compared to the

model's saliency predictions.

Pang *et al.* [PKT+08] proposed a saliency model inspired by Koch and Ullman's theoretical framework on visual attention [KU85]. This model provides saliency estimations based on the predictions of a Bayesian network. This model included a series of novel additions including a model for estimating eye focusing positions and their density distributions. Also the model has the ability to derive the parameters of the Bayesian network using the Expectation Maximisation (EM) algorithm. Another advantage of this model is its computational efficiency. In particular, when data from eye tracking trajectories are accessible, saliency estimations can be obtained almost in real time using GPUs and a Markov chain Monte Carlo particle filter. The results indicate that this model gives better saliency predictions compared to the non-stochastic model proposed by Itti and Koch. The stochastic nature of this model is advantageous as it better approximates the way the HVS system extracts and process visual salient features.

3.5.2 Auditory saliency

Auditory saliency is the term used to describe salient features in the auditory domain. As mentioned in section 3.2.2, auditory attention is the psychophysical process that directs humans to perceive audio stimuli that either contain conspicuous features or because they willingly want to attend to it as part of a task. Auditory saliency models approximate the process of attending acoustically salient information from a complex audio signal that is given as input.

The first model capable of computing auditory saliency was presented by Kayser *et al.* [KPLL05] in 2005. This model simulates the basic processing followed by the HAS in separating sounds from a complex audio stimulus based on spectral and temporal modulation features. Specifically, these modulation features allow to distinguish a sound of interest amongst noise and non-important acoustical information. Similar to the saliency model followed in the visual domain [IKN98], an *Acoustic saliency map* encodes all these audio features that distinguish among all the non-salient audio stimulations. The auditory saliency map incorporates multiple modulation features which, in turn, simulate the processing followed by auditory neurons in extracting interesting audio tokens from an input sound stream.

As far as the implementation of Kayser's model is concerned, the algorithm firstly converts the sound waveform in a time-frequency representation (*Spectogram*) that contains information of how acoustic energy is distributed across the audible range of frequencies (y-axis) and across the time duration of the input audio track (x-axis). An example of a spectogram is given in Figure 3.6. The log transform of the

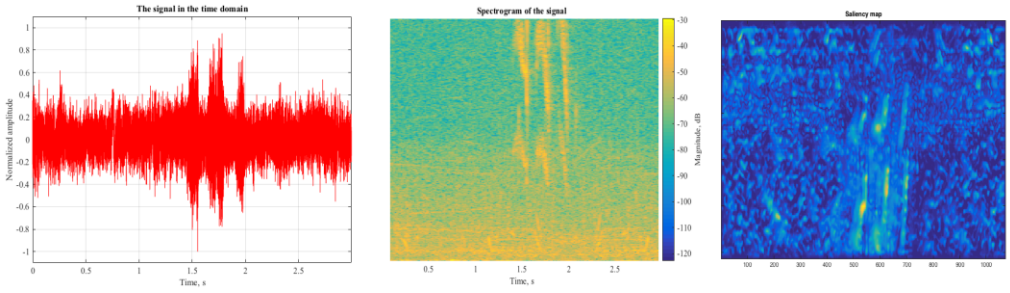


Figure 3.6: **Left:** Waveform representation of a sound signal. Bird chirping at a noisy urban backgound. **Center:** Spectogram of the input audio. Higher acoustic energy is present at higher frequencies where chirping starts. **Right:** Auditory saliency map, salient features are shown with warmer colours. Both the saliency map and the spectogram were implemented following Kayser's model [KPLL05].

absolute values of this image yields an intensity image that enhances the frequencies with higher energy. The second step includes the extraction of multiple features at different scales using the input intensity image from the first step. These features are encoded in images that contain sound intensity, frequency contrast and temporal contrast information and result from filtering the intensity image with a range of linear filters. As was the case in the saliency map in the visual domain, center-surround differences are also used to promote salient features from a cluttered input resulting to a set of feature maps that contain all the acoustically salient features. The linear combination of these feature images gives the acoustic saliency map. An example of an auditory saliency map implemented following Kayser's model is given in Figure 3.6.

Kalinli *et al.* [KN07] used a different approach from Kayser's model in extracting salient features and specifically they used the pitch distribution of the input audio stimulus. According to literature [PUJG02], it is not known whether the HAS extracts pitch information based exclusively on temporal or frequency characteristics of the auditory input. The temporal approach supports the idea that the brain approximates the periodicity of the incoming sound at each auditory nerve using some form of cross-correlation between the auditory stimulus and regions of the auditory input. The authors adopt this approach and use temporal information for understanding how pitch changes over time. This fact allows them to extract salient sound features using a similar methodology like the one followed in the visual domain [IKN98].

3.5.3 Discussion on saliency

Current auditory saliency models tend to be problematic when they are applied to real-life audio recordings. These models operate over extremely short time-scales, even shorter than the typical gaps in a conversation. However, even during these small time windows, background noise is captured and normalised as part of the algorithmic procedure. This normalisation results in the promotion of these "noisy" parts as salient features while in reality they are completely unimportant. Figure 3.6 shows how these noisy artefacts have been marked in the auditory saliency map as scattered orange blobs. It is worth noting, that in the case that a computer generated audio track is given as input the above problem does not happen anymore as the audio is inherently free of background noisy artefacts.

Another point that merits further discussion is the absence of a memory module in the acoustic saliency models. For instance, given a speech audio track, someone can find many salient features when the speaker starts talking but if he/she has been continuously talking for some time, it is easier to start paying less attention making the stimulus less salient over time. The current auditory saliency models do not take this scenario into account and operate at the same manner on many time scales. Another criticism of the current auditory saliency models is the fact that they do not account for sound localisation cues, either binaural or monaural. For example, if a sound is continuously perceived to come from the left hand side and suddenly a novel auditory stimulus comes from a different direction, it is quite possible to be more salient. At the moment, this feature is still an open question to be introduced in existing saliency models.

Checking the validity of auditory saliency models is not an easy task due to the nature of the sound stimuli. In the visual domain, saliency models can be easily validated using eye tracking results from experimental studies and compared with estimations given by the saliency maps. In the auditory domain, ear movement is not possible and any possible head movements cannot directly be attributed to the existence of acoustic salient features. Head movement is usually a typical reaction for localising an audio cue. A naive way to assess auditory saliency would be to evaluate detection of the sound stimulus. If a sound feature is salient then it can get easily detected and memorised. On the other hand, if a feature cannot get detected then there is no point to assess its saliency. Of course this methodology discards all possible intermediate stages where salient features can be identified and extracted.

3.6 Selective rendering

Rendering techniques in computer graphics are used to synthesize a digital image or a RIR from a collection of data defining a virtual environment. This section surveys a class of methods termed *Selective rendering* or *Perceptually-based* rendering techniques that take advantage of known limitations of the human visual/auditory systems for rendering an image or audio in a non-uniform way. The main advantage of these methods is that they can produce visual/auditory stimulations that are perceptually indistinguishable from their full rendered versions with less computational time and resources.

The HVS is able to perceive quality details up to a specific level. This limit sets a threshold on the computational effort required to render a high quality digital image. Specifically, trying to improve visual quality above this threshold is considered an unnecessary waste of time and resources. In order to minimise unnecessary costs, selective rendering methods identify areas that are more likely to be attended by the human observer and favour these regions by rendering them at higher quality compared to the rest of the image. In order to accomplish this task these methods take a series of quality decisions to identify which portions need further elaboration. These decisions are taken either in a pre-computation stage or during the run time of the rendering algorithm.

3.6.1 Visual selective rendering

Yee *et al.* [YPG01] proposed a selective rendering framework based on saliency estimations in order to reduce rendering computational times. Specifically, for every frame a spatio-temporal error map and a saliency map enriched with a motion channel were created. The former was constructed based on velocity dependant contrast sensitivity values and saliency estimations. Both of these maps were constructed using low quality rendered versions of the frame computed either using rasterisation or by computing only the direct illumination component (see section 5.2.2) of the governing rendering equation. The error tolerance map and the saliency map are merged appropriately to yield what the authors term the *Aleph* map. This image guides computational resources in the main rendering step.

In another study, Haber *et al.* [HMYS01] proposed a perceptually-guided methodology for exploring virtual scenes in real time using a corrective splatting algorithm. Before the main computation step, the algorithm approximates a solution to the rendering equation (see 5.2.2) based on tracing particles in the scene. During the main computation step, the remaining computational resources are allocated to

improving the quality of non-diffuse regions in the virtual scene. These important regions are identified through the use of a saliency map computed from the low quality image that obtained in the pre-computation step. The authors use Itti and Koch's saliency model [IKN98] in their implementation.

Cater *et al.* [CCL02, Cat04] demonstrated in practice how the phenomenon of Inattentional Blindness (see section 3.2) can be exploited to speed up computation times when rendering animations. In their framework, all areas related to an assigned task were rendered at higher quality whilst the quality at the non-relevant regions was rendered significantly lower. The authors conducted psychophysical experiments and validated that subjects who performed the task, failed to perceive the low quality areas of the frames that composed the animation. This study clearly demonstrated that the top-down visual attention mechanism can be successfully utilised in the development of selective rendering frameworks.

In a related study by Cater *et al.* [CCW03], the authors achieve to imitate the HVS's property to ignore task irrelevant objects and encode them in what they term as a *Task map*. Specifically, in their framework, a first order rendering of the frame is computed which serves as the starting point to extract object and motion related information. The extracted data are combined with information from a data structure that compactly encodes the importance of each object in the scene and yields the task map. The authors make use of the task map along with the draft rendered version of the image and a contrast sensitivity model to render every frame of the animation in a progressive manner. Their results show significant speed-up compared to classic rendering techniques.

Sundstedt *et al.* [SDLC05] also used the idea of the task map as was defined by Cater *et al.* [CCW03] and that of the saliency map from Itti *et al.* [IKN98] in order to construct what the authors name as *Importance map*. This map applies different weighting coefficients to the task map and the saliency map in order to obtain combined information of both top-down and bottom-up visual attention characteristics. Specifically, the importance map is denoted as $\mathrm{IM}(\omega_T, \omega_S)$, where ω_T, ω_S are the weighting coefficients applied to the task map and saliency map respectively. For instance, a saliency map can be defined as $\mathrm{IM}(0, 1)$ while a task map as $\mathrm{IM}(1, 0)$. An importance map that contains equal contributions from the other two maps can be written as $\mathrm{IM}(\frac{1}{2}, \frac{1}{2})$. This work demonstrated that importance maps can be used in rendering animations at Selective Quality (SQ) with no perceptual difference compared to the uniformly computed versions of the images at High Quality (HQ).

Another form of selective rendering, as proposed by Debattista and Chalmers

in 2005 [DC05, Deb06] included the application of path tracing for computing a global illumination solution based on adaptive sampling on the components of the scene. This novel methodology was termed *Component-based adaptive sampling*. Specifically, earlier work used adaptive sampling at pixel level to speed-up the rendering times. This sampling methodology calculates radiance differences at pixels and estimates areas of high variance where more rays need to be shot compared to the rest of the scene. The authors applied the adaptive sampling technique on the reflectance properties of the various objects that composed the scene instead at pixel level and showed a significant speed-up at the rendering times compared to the previously used adaptive methods.

Longhurst *et al.* [LDC06] developed a GPU saliency model for selective rendering applications. Their work included a series of novel contributions related to the selective rendering methodology and the estimation of the visual saliency itself. Primarily, it was implemented on a GPU giving results around seventy times faster than the conventional CPU implementation. The authors concentrated on identifying salient features at every pixel composing the image instead of discovering which region was attended first. In the saliency model, the orientation channel was replaced by an edge detector while the implementation also included a motion channel as was proposed by Yee *et al.* [YPG01]. Furthermore, the authors added a *habituation* channel to account for the effect of the HVS's short-memory feature. This habituation module preserves the saliency of a virtual object for some time depending on its appearance in successive frames of an animation and after that it suppresses its saliency for about ten seconds to simulate the effect of temporal memory.

3.6.2 Perceptual metrics for images

Metrics have been proposed in the past for estimating perceptual differences between two images. These metrics take into account the properties of the HVS and imitate its ability to distinguish visual features that are not unique to a pair of images. A well known metric for this task is the Visible Differences Predictor (VDP) introduced by Scott Daly in 1993 [Dal92]. This metric gets as input a pair of images and outputs a probability map that shows which regions are more likely to trigger a perceptual difference to a human viewer. VDP operates at pixel granularity and is composed of several modules that operate similar to the HVS's inherent behaviour to react non-linearly to changes in luminance and contrast. These perceptual characteristics of the HVS are taken into account at every stage of the model using properly scaled psychometric functions.

The proposed VDP model takes the HVS perception to input light stimuli into account and uses several perception based models such as *Amplitude compression*, *Contrast detection*, *Cortex transform* and *Visual masking* to predict perceptual differences between input pairs of images. Figure 3.7 gives a graphical representation of the different modules described in the VDP model [Dal92]. The figure depicts the flow of data for obtaining the perceptual differences encoded in the output probability map.

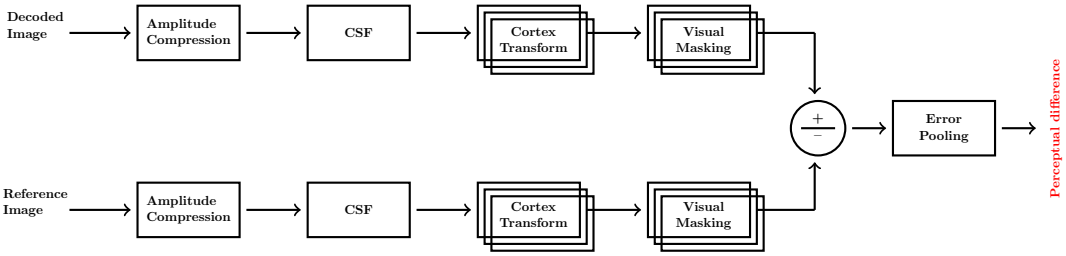


Figure 3.7: The main modules of the visual differences predictor computational model as described in Daly's study [Dal92]. Two streams of data are used for a reference and a test image. The main output is a colour map that shows the probability a human viewer to perceive a difference at different regions of the image.

The first module in the pipeline (amplitude compression) accounts for the non-linear response of the HVS to input stimuli. In the second module (contrast detection), the image is weighted by a contrast sensitivity function to account for the non sensitivity of the HVS to very high or very low spatial frequencies in the image. The last two modules (cortex transform and visual masking) break down the image into a spatial and an orientation channel and estimate perceptual differences in each channel independently.

The HDR-VDP proposed by Mantiuk *et al.* [MDMS05] is a dedicated HDR extension to the original VDP model. The primary change between the VDP and the proposed extension is that the former assumes a global adaptation to luminance whereas the latter assumes that viewers can adapt to every single pixel of the target image. This assumption leads to a conservative estimate of the image quality as well as introduces more reliability to the quality prediction of the proposed metric. Unlike the VDP which models the photo-receptor response to input stimuli, the HDR-VDP uses a perceptual uniform JND scale to non-linearly transform input physical luminance values to perceptually uniform code values. Also unlike the original VDP where the CSF model was responsible for modelling the loss of sensitivity and normalisation of contrasts to JND units, the CSF model in HDR-VDP does not

require any normalisation since luminance values are already scaled to JND units.

The original HDR-VDP was further improved in a later study by Mantiuk *et al.* [MKRH11] and a new extension proposed (HDR-VDP-2). The primary contributions were firstly, a generalisation of a broad range of viewing conditions. Secondly, the proposal of a comprehensive visual model derived from a psychological evaluation which takes into account several properties of the HVS such as intra-ocular light scatter, photo-receptor spectral sensitivity, separate rod and cone pathways, intra- and inter-channel contrast masking and spatial integration. Thirdly, the revised model includes improvement of the supra-threshold quality metric predictions (i.e. the distortions are clearly visible to the human eye). However, a criticism of HDR-VDP-2 was that the error prediction was accomplished by the pooling of errors in several frequency bands where the weights were determined by optimising an existing LDR dataset which limits the accuracy of the prediction in high contrast scenes. Also, the optimisation was performed on a relatively small set of images and was unconstraint which led to negative pooling weights which were not easy to interpret.

Narwaria *et al.* [NMDSLC15] proposed an extension to HDR-VDP-2, namely, HDR-VDP-2.2 which addressed the issues with error pooling. The pooling weights were re-optimised on a combined dataset of LDR and HDR images which resulted in more effective predictions for both LDR and HDR test conditions. Moreover, the optimisation was formulated to be constrained such that the resultant weights can be computed in a bounded way. The work presented in chapter 5 of this thesis makes use of the latest optimised version of HDR-VDP on pairs of LDR images.

3.6.3 Auditory selective rendering

Similar to the visual domain, researchers in the acoustic domain exploit limitations of the HAS in order to look for possible optimisations when rendering audible stimuli. In acoustics, the main computational costs are attributed to the audio spatialisation operations needed for every one of the sound sources present in the virtual scene. In cases where multiple sounds are present the HAS utilises what is known as *Auditory masking* for ignoring perceptually unimportant auditory stimulations reaching higher levels of processing [Moo03]. Auditory masking is the cornerstone of the *Perceptual Audio Coding* (PAC) that is used in advanced audio compression schemas. The well known MPEG 1 Layer (mp3) standard [BP00] makes use of auditory masking to compress audio streams without the listener being able to understand any perceptual difference between the compressed and uncompressed audio streams.

Moeck *et al.* [MBT+07] used saliency estimations extracted from an acoustic saliency map [KPLL05] in order to compute the acoustic features that attract human auditory attention at higher quality. In their method, the authors render audio by integrating saliency values over frequency regions (subbands) while they use saliency estimations as a mean of clustering multiple audio sources. Similar methodology for grouping audio sources using perceptual criteria, was also followed by other researchers [GLT05, Tsi05] and includes phase composition of the audio signal by combining its instantaneous acoustic energy with the energy during its attenuation stage.

Tsingos *et al.* [TGD04] used Level-Of-Detail (LOD) and auditory masking in order to facilitate auditory rendering in the spatial domain when hundreds of auditory sources are present in the virtual scene. Specifically, the authors use sound loudness as a metric of importance and accumulate the importance of a finite amount of sources until their loudness masks the importance of all the remaining sound sources combined. Every time the algorithm finds a sound source that represents a set of other sources it is placed in the centre of the group and is adapted over time based on the importance of the other sources in the group. As far as the implementation is concerned, authors combine the sound sources using a k-means clustering algorithm.

3.6.4 Cross-modal selective rendering

Cross-modal selective rendering frameworks exploit perceptual interactions between the senses in order to selectively render one of the involved modalities in lower quality without any perceived quality artefacts. The effect of a sensory stimulation on another modality, termed the target modality, can be considered either in the *spatial* or the *temporal* domain. For instance, the effect of auditory cues on perceiving visual stimuli can be exploited in rendering either static images or sequences of frames. This section, firstly, surveys how rendering visuals can be accelerated under the effect of auditory or olfactory stimulations at the same time and secondly, how audio rendering times can be significantly decreased under the effect of visual cues.

Rendering visuals under the effect of auditory stimuli

Storms *et al.* [Sto98] showed that the level of realism in a bi-modal virtual environment depends on the quality of both of its components, namely, visual and auditory. Specifically, the authors demonstrate that the perceptual quality of a HQ video can be further increased when coupled with HQ audio. On the other hand, the study

shows that HQ video can significantly reduce the perceived quality of a Low Quality (LQ) auditory stimulus.

Mastoropoulou *et al.* [MDCT05a, Mas05] showed that the presence of a Sound Emitting Object (SEO) in a virtual scene can be used to selectively render animations at significantly lower rendering times. In their study, the visual region contained the SEO was rendered in HQ at an appropriate angle, while LQ visuals were displayed at the rest of the scene. This study exploited the angular sensitivity of the HVS while the phenomenon of Inattentional Blindness was the perceptual mechanism that directed viewers' gaze at the active SEO leaving large portions of the scene unattended. An extrapolation of this study by Harvey *et al.* [HWBR$^+$10] included the application of spatialised sound in a VE context. They showed via eye tracking that visual attention is significantly affected by the direction of the incoming audio stimulus.

In a recent study by Harvey *et al.* [HDBRC16], the effect of spatialised sound on visual attention is examined through the combined use of both auditory and visual saliency estimations. The combined information is encoded in what the authors term as a *Multi-modal* map. Specifically, based on a sound transport simulation in a VE they construct a sound map that provides saliency information regarding the direction of the auditory stimulus. The method for constructing the sound map is based on tracing phonons in the scene and was proposed earlier by Bertram *et al.* [BDM$^+$05]. The authors make use of the multi-modal map to effectively reduce the rendering times in a series of scenarios while they validate its effectiveness through an experimental study where participants fail, on average, to perceive quality artefacts at the displayed images.

In another study, Mastoropoulou *et al.* [MC04] investigated through an experimental study whether audio stimulations can affect visual temporal perception. The authors used two different types of music including slow tempo relaxing music and fast tempo impelling music delivered together with an animation. Both of these conditions were compared with a control condition where no auditory stimulus was delivered to the user. The results indicated that there was no significant effect of the two types of music on the perceived frame rate of the delivered animations. This outcome contradicts with the auditory driving phenomenon and the modality appropriateness hypothesis, discussed in section 3.3. The authors explained this outcome as follows: firstly, the frame rate of the animation was only 4 FPS which was assumed to be quite low and secondly, the duration of the animation was 40 seconds which is much higher than HVS's short-term memory capacity. According to Atkinson and Shiffrin [AS71], short-term memory lasts for about 15 to 30 seconds.

In a similar study by Mastoropoulou *et al.* [MDCT05b], the authors investigated the effect of congruent or irrelevant sound stimuli on the perceived quality of an animation. In their experimental study they used eye tracking to compare the saccadic movements and fixation times at two different cases. The first case included watching an animation coupled with the congruent/incongruent sound stimulus while the second included watching an animation at high frame rate with no accompanying sound stimuli. The results showed that the congruent sound stimulus attracted participants' attention and allowed the frame rate to be decreased without the observer being perceptually affected by the lower delivery rate of the frames .

In a study by Hulusic *et al.* [HCD+09], the authors carried out research on the correlation between audio beat and video frame rate on static or dynamic animated sequences. Specifically the authors investigated the effect of beat rate, scene type (static/dynamic) and subjects' previous familiarity on the perceived frame rate of the animation. The results indicated that beat rate and frame rate are closely connected at multi-sensory virtual environments. In the case of static sequences, low beat rates had a considerable effect on the perception of low frame rates while the results showed that previous familiarity had no effect on the perceived frame rate.

Hulusic *et al.* [HDAC10, HDAC11] investigated the effect of movement related sound effects on watching animations. In this study, the authors considered the effect of including the sound effect of footsteps on walking related animations. They found that the sound stimulus increased the perceived sleekness of the animation. Specifically, the results showed that animations delivered at 20 FPS with congruent sound effects were evaluated as significantly smoother than animations played without sound stimulus at frame rate either 30 FPS or 60 FPS. It is worth pointing out that the authors attempted to find a correlation between movement sound effects with running related animations with no success. This outcome indicates that the scenario content matters when coupling with audio or participants were able to discriminate synchronisation problems between audio and video.

Rendering visuals under the effect of olfactory stimuli

Brkic *et al.* [RCHR07] exploited the presence of olfactory stimulations to direct participants' visual attention towards a source of smell or smell related objects in the virtual scene. In their methodology an image was selectively rendered at high quality at the smell source locations while lower quality was applied at the rest of the scene. This methodology was assessed using an experimental study where, on average, participants were not able to identify quality degradations between the

SQ image and its uniformly rendered version. It should be noted that in this study, authors did not use congruent smell-image pairs while the delivered smell was generic and not related to the delivered visual cues.

Brkic *et al.* [BCB+09] exploited previously known knowledge of cross-modal interactions between vision and olfaction in order to investigate the effect of lawn grass smell on the perception of animations that represented grass fields. The smell of cut grass or CIS-3-hexenol, was presented using a commercial atomiser whilst the visual input included two animations of a grass terrain, one was rendered in HQ using various global illumination features (shadows, diffuse interreflections, etc) whereas the second was rendered at LQ with no such features. The results showed that around 80% of the participants were capable of telling the difference between the two presented animations but around half of them could not find any perceptual differences when odours where delivered at the same time as the visuals.

In another study by Brkic *et al.* [BC10], shows how the presence of smell, sound and temperature gradients can impact the perceived quality of a rendered animation. Their experimental results indicated that promising computational savings can be achieved with the presence of diverse sensory stimulations in virtual environments. This study also showed that different sensory level combinations can be more influential on the perception of visual stimuli. Specifically, high values of smell concentration, sound loudness and temperature were more effective than other level combinations of the three aforementioned senses.

Harvey *et al.* [HBRDC11] exploited the effect of olfactory attention on rendering images in a selective manner. Specifically, their study included the implementation of two closely related experiments. In the first experiment, eye tracking data were collected using a series of images that were presented to subjects under the effect of a congruent smell stimulus or no odour at all. The eye attention points and fixation times validated that when odour was delivered, vision was directed towards smell-related objects. The authors modified the Itti and Koch's saliency model [IKN98] implementation to take into account the eye-tracking data. The new model can predict salient features based also on olfactory information. The modified saliency map was used to selectively render images that were found perceptually indistinguishable, on average, from the full rendered versions in a second experimental study. Although, this study showed a generic way to exploit olfactory stimuli on selective rendering frameworks it does not quantify the smell concentration needed for obtaining the effects of odour in the saliency map. For instance, faint smells could have lower or no salient features therefore the saliency map should be adjusted accordingly.

Rendering acoustics under the effect of visual stimuli

Selective rendering of audio under the effect of visual or olfactory stimulations has mainly remained unexploited in multi-sensory virtual environments. This section surveys studies that represent this direction. Grelaud *et al.* [GBW+09] considers audio rendering when multiple visual objects collide with each other. The computational requirements for spatialising sound effects for every individual sound source is prohibitive and the authors considered a sound synthesis approach in combination with pre-recorded audio effects. In their method the instantaneous energy of the emitted sound signal is precomputed using the recorded audio whilst the effects of multiple object collisions were simulated using a quick physics estimation (sand box). This study demonstrated that audio Level Of Detail (LOD) can be increased while reducing visual details. The end result remains perceptually the same for the viewers.

Discussion on cross-modal selective rendering

The constant need for high quality VEs capable of immersing the user will possibly increase the presentation of multiple and diverse sensory stimuli in the near future. Previous work in the field of cross-modal selective rendering shows that these techniques have potential for reducing the computational times while keeping a satisfactory level of fidelity for every delivered sensory stimulus. Although many sensory interactions have been utilised in selective rendering frameworks there are known perceptual phenomena that still remain uninvestigated in the area of computer graphics.

Cross-modal frameworks have mainly targeted the effect of one sense on the perception of the other, exploiting the perceptual interactions of auditory/olfactory stimulations on computing and delivering visual cues at selective quality. In real life scenarios, people receive sensory information from diverse sources, and adjustments in one sensory modality can perceptually impact all the involved senses to a certain degree rather than a small subset of them. This characteristic has not yet quantified and used for reducing computational requirements in multiple senses simultaneously within a VE framework. A unified methodology that would allow to study the degree one sense can be perceptually affected by the presence of others could enable the derivation of a metric for evaluating the importance of every sense in a multi-sensory environment.

All the cross-modal methods discussed in this chapter have explicitly looked into the required fidelity per sense as computational resources change. Techniques

have investigated the benefit of resource fluctuation within a multi-sensory context, based upon psychological evidence for the interactions between the senses but, to the best of author's knowledge, no previous work has looked into strictly quantifying this process.

3.7 Summary

This chapter presented an overview of a growing area of computer graphics known as *perceptually based rendering*. Basic anatomical characteristics of the human sensory organs were reviewed first. The ability of these organs to selectively attend specific areas for further processing enabled the development of selective rendering methods for significantly reducing rendering times with no perceptual impact on the viewer. All the sensory systems have certain spatial and temporal properties which are the primary reason for perceptual interactions between the senses. These interactions have been exploited in multi-sensory VE frameworks for reducing rendering times of one of the involved modalities without any artefacts to be consciously perceived.

While interactions between the senses have been previously used to reduce the computational requirements, none of the selective rendering frameworks presented in this chapter has formally investigated into how the process of allocating computational resources can be quantified formally based on human selection criteria. The importance of each sense based on the context of the scene and the availability of the resources has not, to our knowledge, been investigated in previous studies. According to previous experimental hypotheses by Posner *et al.* [PNK76], visual quality is always considered crucial for the VE users although there is no evidence under what conditions (budget/scenario selection) this is the case. Furthermore, the validity of the hypothesis when olfactory stimuli are delivered along with visual and auditory cues also needs to be examined through experimental studies in tri-modal VEs.

Chapter 4

Research methodology

4.1 Introduction

Previous work, as described in chapter 3, has shown that humans are not able to fully attend to all the incoming sensory stimuli at the same time. While many limitations of the Human Sensory System (HSS) have been previously used to reduce computational requirements [HHD+12, SM93], precisely how best to allocate a computational budget in the simulation of multiple sensory stimuli has, to the best of our knowledge, not been investigated before.

The methodology presented in this chapter addresses the question of how the perceived fidelity of a multi-sensory experience is affected by the amount of available computational resources given for rendering the sensory stimuli. The goal is to assess the quality of the sensory cues that comprise the VE as a function of the available computational budget based on people's preferences. These preferences are captured in an experimental study and are used in the construction of a prediction model capable of estimating the percentage of the total resources that need to be devoted for obtaining better quality stimuli in each of the senses present in the VE.

The research methodology firstly approaches the problem in the context of two senses, namely, sight and hearing and studies how people tend to distribute a given amount of resources for obtaining better quality visual and auditory stimuli. Human allocation preferences are captured in an experimental study (***E1***) where the availability of resources ("budget") and the scenario change. The majority of the previous research studies utilised perceptual interactions between the senses of sight and hearing as these are the most commonly used senses in VE set-ups, thus this thesis considers the case of bi-modal VEs with audio and visual cues as a starting point.

Following the two senses experimental study, the research methodology considers how multiple senses can be introduced in the experimental framework for the objective of resource allocation. The importance of different sensory stimuli in the same framework can be very different depending on the available resources and the scenario displayed to the users. As long as there is the software and hardware infrastructure to render and deliver these stimuli, the methodology described in this chapter can be used as a guidance for any number of different sensory cues for distributing resources based on human preference criteria. In order to demostrate this method, physically accurate smell impulses are simulated and delivered along with the audio-visual stimuli in a subsequent experimental study (***E3***).

The simulation and delivery of olfactory cues significantly differs from the senses of vision and hearing and there is no standard way to quantify quality of olfactory stimuli. Smells are evaluated in a subjective manner by humans based on previous experiences and familiarity with the exposed smell [HCL03, PSNG06]. Perceptual properties regarding the intensity of a delivered odour are examined in this thesis based on a psychophysics experimental study (***E2***) that aims to investigate the JND threshold of a reference smell.

This chapter is organised as follows. Firstly, related work that presents similar experimental methodology is presented. Secondly, the experimental method for answering the research question when two senses are present in the VE is considered and analysed. Thirdly, the introduction of more senses in the experimental framework is discussed. This thesis moves towards the inclusion of the sense of olfaction, although, the same methodology can be used for the introduction of any other sense in the resource allocation framework. Finally, having considered all the olfactory related challenges, an experimental method that incorporates all the three aforementioned senses (sight, audio and olfactory) in the same experimental framework is proposed and exploited for distributing computational resources in multi-sensory VEs. Figure 4.1 shows pictorially how the experimental studies ***E1***, ***E2*** and ***E3*** are organised in the chapters of this thesis.

4.2 Related research methodology

The relative importance of audio and visual stimuli has been investigated when assessing the quality of multimedia applications. In a study by Pinson *et al.* [PIW11] the authors reviewed a series of experiments that were used to evaluate the importance of each of the two senses in a range of different source materials, video resolutions, audio and video compression schemas. Each of these experimental studies

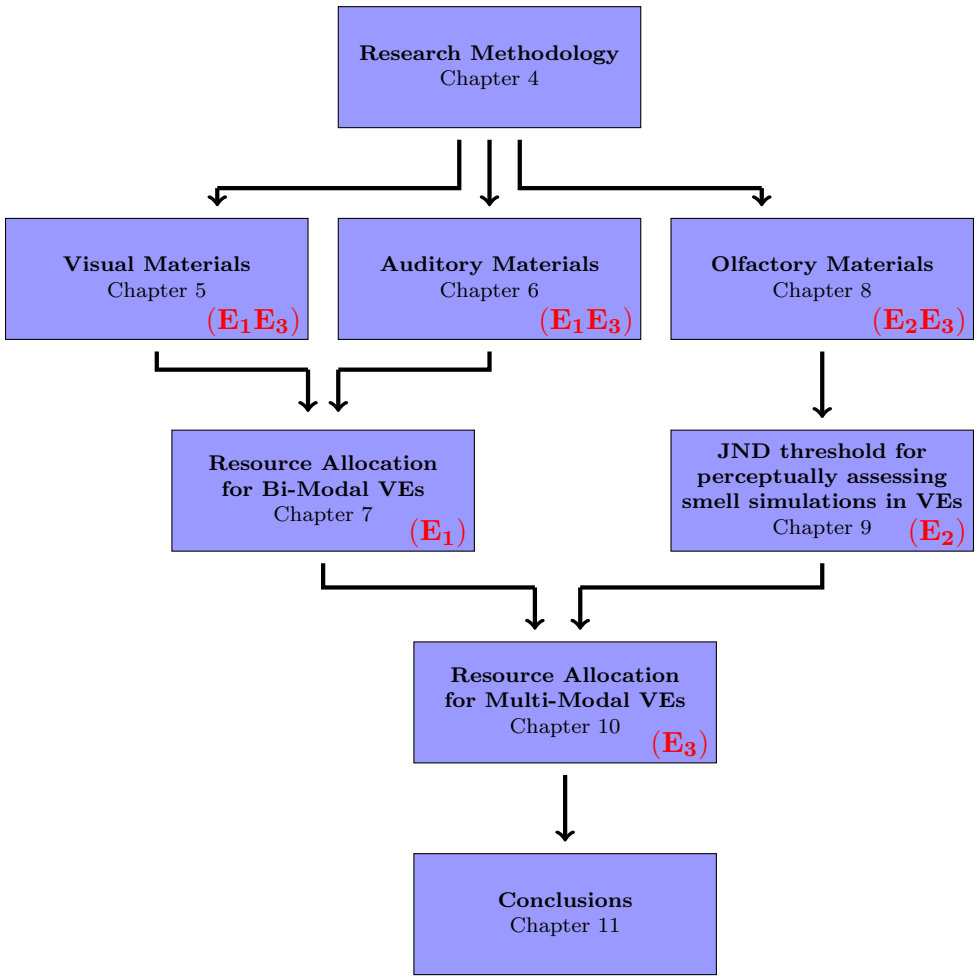


Figure 4.1: Schematic structure of the contents of this thesis. Every chapter is served by the contents of the chapter above it in the diagram and serves the chapter below it. The red letters in parenthesis are the notations used for every experimental study.

produced a model that estimates audio-visual quality based on the measured visual, auditory and combined audio-visual Mean Opinion Scores (MOS).

In another study by Belmudez and Möller [BM13], the effect of audiovisual quality is investigated in the context of interactive communication services. The authors test the effect of cross-modal interactions in a range of different experimental conditions where the participants have either a passive or active role in the communication context. This work proposes an integration function for predicting audio-visual quality assessment depending on the type of communication application.

Although these methods investigate the relative importance of one sense over the other, they do it for assessing the quality of multimedia applications and not for efficient resource allocation in the context of VEs. Furthermore, these studies do not consider how the audio-visual importance relation changes when the amount of available resources varies as they do not take into account the computation effort needed for rendering the sensory stimuli. Methodologies described in this section and the literature review (see section 3.6) focus on the benefit of resource savings due to the presence of multiple senses in the VE context but they do not explicitly look into quantifying this process. Allocation of the available resources at the two senses depending on user's fidelity preferences has not been investigated, to the best of our knowledge, in the design and delivery of virtual experiences.

4.3 Research method

The research method is described first in the context of two senses. Considering resource allocation using bi-modal interactions allows better understanding of the experimental framework and facilitates the subsequent introduction of other sensory modalities. The use of visual and audio cues in bi-modal VEs is quite common and the majority of the research studies presented in chapter 3 make use of these senses. Since olfaction presentation is more primitive that audio and visual stimuli, a significant amount of work was conducted to enable the simulation and delivery of olfactory impulses. The intensity of olfatory stimuli is also examined at perceptual level to investigate differences between the range of available concentrations that can be delivered by our available olfactory display and are perceptually distinguishable to the users. Finally, this section concludes by describing the research methodology of an experimental study that encompasses all the three senses and builds upon the experimental layout and observations obtained from the previous two experimental methods.

4.3.1 Research methodology for audio-visual interactions

The objective of this research is to study user allocation preferences when distributing visual and audio resources while the available computational budget varies. Efficient schemas for carefully allocating computational resources are of significant importance both in industry and academia as they allow the rendering process to be adjusted depending on the importance of each of the delivered senses at any given time.

The initial hypothesis is that high quality imagery is often appreciated more,

compared to audio, when experiencing a virtual world. This hypothesis is investigated through an experimental study. In this study, resource allocation is measured over a series of experimental scenarios and budget sizes whereby participants are asked to assign the given budget to experience better audio or visual cues. Therefore, the effect of two independent variables, namely, experimental scenario and budget size, are considered on the percentage of the total budget that is assigned for visual quality. This percentage is the dependent variable of the experiment ***E1***. The budget percentage given to update the audio quality can be deterministically given by the percentage devoted to graphics.

The objective of this experiment is twofold. Firstly, to investigate how the visual-audio importance relationship scales across different budget sizes that can be available to the user and secondly, to utilise the obtained experimental data for building a prediction model capable of estimating resource allocation depending on an input budget size and scenario.

The experiment was designed using rendered images and audio where the quality level of each of the two senses could be adjusted interactively using the controls of a Graphical User Interface (GUI). Physically-based simulations were used for the computation of both audio and visual stimuli. In order to make use of the obtained experimental data in actual applications, a statistical model was designed. This model takes the computational budget and scenario as inputs and provides an audio-visual ratio estimation. A second, complementary experiment, was conducted with a new set of participants and untested budgets to validate the model's performance against actual participant selection preferences.

The process of rendering visual and auditory stimuli is described in chapters 5 and 6 (see also Figure 4.1) of this thesis along with the rendering background required for the explanation of the computation process followed in either of the two modalities. The description is separated in two different chapters for better clarity. Chapters 5 and 6 also define the metrics used for assessing quality for each of the two considered senses. The selected quality metrics are associated with cost functions for estimating the computational cost required for obtaining better quality levels in vision and audio. Before implementing the experimental study ***E1***, it was important to have a predefined set of stimuli that perceptually differed from each other at each of the two considered senses. All the audio-visual stimuli needed to be presented in perceptually uniform steps that could elicit distinct visual/auditory differences and would have different computational costs. In both vision and hearing there are methodologies to evaluate whether two stimuli are perceptually different either in the form of a computational model or using previously estimated JND thresholds

for some characteristic of the considered stimulus (e.g. contrast or brightness in vision and loudness or pitch in acoustics).

4.3.2 Introducing multiple senses in the experimental framework

As explained in the previous chapters (see section 3.6), previous research has concentrated mostly on exploiting perceptual interactions between the senses of sight and hearing to selectively render VEs with lower computational cost. The introduction of other sensory modalities has remained relatively unexplored in this area. Many senses in a VE can significantly increase the level of immersion and give a more genuine representation of real life scenarios. Efficient resource allocation schemas in multi-sensory VEs is more a necessity rather than a requirement as rendering all these sensory stimuli in high quality is considered computationally prohibitive and offers no perceptual benefit to the users.

The introduction of more senses in the resource allocation framework includes challenges both at a technical and perceptual level. A technical challenge has to do with delivering physically correct stimuli intensities to the users synchronously with other defined modalities using specialised hardware. On perceptual level, every sense has its own properties that are required for identifying distinct differences between delivered stimuli of increasing computational cost. Measuring quality for senses other than vision and hearing might not be as straightforward as quality is mainly evaluated subjectively and not objectively by the users. Also, there are no well defined standards that can assist a metric selection. In this thesis, extension of the resource allocation framework from bi-modal to tri-modal VEs is achieved by considering the sense of olfaction along with the senses of sight and hearing that were introduced earlier.

The sense of smell was included for two reasons. Firstly, smell intensity estimation can be achieved through the simulation of smell transport in a VE via Computational Fluid Dynamics (CFD). CFD techniques simulate the physical processes that govern smell transport and estimate odour concentration at any point of the computational domain. Physically accurate concentration values are then directed to an olfactory display for delivery to the VE users. Smell transport simulation and delivery issues are addressed in more detail in chapter 8 of this thesis. Secondly, the availability of an olfactory display assisted to estimate olfactory perceptual thresholds and search for an olfactory quality metric before introducing the possibility of delivering smell impulses in the resource allocation framework.

There is a lot of research in the visual and auditory domains regarding whether two given stimuli (images or sound tracks) are perceived perceptually dif-

ferent for the average viewer/listener [MDMS05, Gel09, JJS93]. In the olfactory domain, although there are references that point out the importance of the sense of smell in real life environments [HE96, YN06], the knowledge of the HOS's perceptual characteristics is limited compared to the visual and auditory sensory systems [BMVK14]. Perceptual thresholds for many odorous compounds have been estimated in previous work [SW55, MC80] focusing mostly in estimating detection thresholds and not JND thresholds.

The work presented in chapter 9 (see also Figure 4.1) examines whether numerical accuracy of an odour transport simulation has any perceptual impact on the perceived smell intensity sensed by the users of a VE. The hypothesis that better numerical accuracy in a CFD simulation does not elicit perceptual differences to the users is examined through the implementation of a psychophysics experiment for investigating the JND intensity threshold for a reference smell. Specifically, in order to identify discernible differences, the numerical accuracy of the CFD simulation is adjusted using finer spatial discretisation of the geometrical domain thus requiring further computational resources (see chapter 8). The selected tool for checking whether better spatial discretisation is necessary is the JND threshold. This is estimated for a range of concentration values via pairwise comparisons of smell stimuli delivered to the participants of the experiment using an olfactory display.

4.3.3 Research methodology for audio-visual-olfactory interactions

The experimental study ***E2*** extrapolates the resource allocation methodology to tri-modal VEs. The initial hypothesis suggests that better spatial discretisation refinements of the computational domain in smell transport simulations do not elicit any perceptual difference to the average user. This hypothesis is experimentally validated with four experimental scenarios.

According to Jiang *et al.* [JCT95], odour intensity cannot be considered a valid measure of smell quality because higher concentration olfactory impulses do not necessarily mean better quality. This study argues that odour intensity can be classified to a scale system at which very intense smell stimuli are described as "intolerable" when people are sniffing them. This result is also intuitively understood from every day experience where very strong odours are not favoured by people despite how pleasant the smell is. Another limiting factor is also the fact that humans evaluate olfactory stimulations in a highly subjective manner affected mostly on previous experiences and familiarity with the exposed odour. This behaviour is called hedonic assessment and has been described in detail in section 2.4.2.

The absence of a metric that can be used to measure smell quality is a limita-

tion, as allocating resources is based on the quality of the stimulus that is delivered to the user and better quality stimuli should have higher cost allocated from the user's total budget. As odour intensity does not imply better quality, the experimental framework of the tri-modal experimental study builds on the same methodology proposed in ***E1*** for visuals and auditory and it extends it by introducing the ability to deliver olfactory impulses to the user in an ON/OFF fashion depending on his/her preferences. When the user "turns" the control for smell ON, a one second smell burst is delivered to the nose using an olfactory display device.

The methodology of the third experimental study (***E3***), considers the effect of two independent variables, namely, scenario and budget size on three dependent variables. These are the percentage of the total budget devoted for graphics and audio quality (numeric variables) and the decision to receive physically accurate smell impuses (categorical variable) when the related control is enabled. The results of this experimental study are analytically given in chapter 10.

As is the case with experiment ***E1***, a prediction model is proposed that can be used for allocating resources in tri-modal VEs. This model is designed using the collected experimental data and validated against participants' selection preferences in a subsequent experimental study, denoted as $\boldsymbol{E3_V}$. The results of the validation stage are also given in chapter 10.

4.4 Summary

This chapter describes the research methodology of the experimental studies presented in this thesis. The implementation of a resource allocation experiment using two senses is considered first. This framework includes the delivery of audio-visual cues where the key idea is to allocate a given budget of resources depending on the required stimulus fidelity in each of the senses of vision and hearing. This experiment was denoted as ***E1***. The process of evaluating numerical accuracy of smell transport simulations using human perceptual criteria is introduced next. The estimation of the JND threshold for a reference odour is used to assess whether high spatial discretisation of the VE is necessary in smell dispersion simulation problems. This study is denoted as ***E2*** in this thesis. Finally, the sense of olfaction is considered along with visual and auditory stimuli in a multi-sensory VE setup. Experiment ***E3*** describes the implementation of a resource allocation experimental study where smell impulses are delivered to the user in a binary manner depending on a user's choice to select the smell ON or OFF when distributing the given budget to the three modalities.

Chapter 5

Graphics pipeline and materials

5.1 Introduction

This chapter explains the process of computing and perceptually assessing the digital images of the experimental studies ***E1*** and ***E3***. In the first part of the chapter (section 5.2), the graphics rendering background for computing the images is given. The computation of the visual stimuli was based in physical-based simulation for obtaining realistic illumination effects. In the second part of the chapter (section 5.3), the quality metric for assessing the rendered images is described. This metric is used to associate the visual images with a theoretical cost that increases when the quality of the image is higher. Finally, the VDP computation model is utilised to find the images that are perceptually similar to the users, on average, but have different theoretical costs. The subset of the perceptually distinguishable images is used for implementing the experiments of this thesis.

5.2 Graphics rendering background

The simulation of light transport in an environment depends on the scope of the application and the physical phenomena that are to be simulated. *Quantum Optics* and *Wave modelling* are theories that explain and analyse the behaviour of light depending on its nature as a bundle of photons or a wave respectively. For the purpose of generating digital images in computer graphics a simpler model is used, known as *Geometric Optics*, which assumes that light is transmitted in straight lines (rays) of infinite speed in an environment. The basic assumption of this model is that the wavelength of light is very small compared to the size of the surfaces that the light interacts with. This assumption allows the simulation of the wave

phenomena described in section 2.3. In this chapter, the terminology and basic definitions from [DBB06] are closely followed for the description of the methods used for rendering images.

5.2.1 Radiometric quantities

All the radiometric quantities are based on the *Radiant power* or flux, denoted as Φ, and measured in Watts (or Joules/sec). This physical quantity expresses how much light energy is emitted/arrived or passed through a surface per unit time. Specifically, when light energy arrives at a surface of known area A, *Irradiance* is used to express the amount of radiant power arrived per unit area (units: Watts/m^2). Irradiance is denoted with E and is given as:

$$E = \frac{d\Phi}{dA}. \tag{5.1}$$

In the same way, when light energy originates from a light source of some area A, the term *Radiosity* is used to define the light energy that is emitted per unit area using the same formula for Irradiance. Radiosity is denoted with the letter B.

Finally, *Radiance*, denoted as L, expresses the power flux per unit projected area and per unit solid angle and is measured in Watts/(steradians $\cdot$ m^2). Steradians is the SI unit of solid angles. Solid angles are explained in more detail in Appendix A. Radiance can be expressed as:

$$L(\mathbf{x}, \Theta) = \frac{\mathrm{d}\Phi}{\mathrm{d}\omega \mathrm{d}A \cos\phi}, \tag{5.2}$$

where $\mathrm{d}\omega$ is the differential solid angle subtended at a point $\mathbf{x}$ which is located at the center of a differential area $\mathrm{d}A$ (see figure 5.1). The point $\mathbf{x}$ is surrounded by a unit hemisphere which defines all possible incoming light directions. These directions can be given for different values of azimuth ($0 \leq \theta \leq 2\pi$) and elevation ($0 \leq \phi \leq \frac{\pi}{2}$) as a pair $\Theta = (\theta, \phi)$. It is usual to denote radiance that leaves a point $\mathbf{x}$ to some direction Θ as $L(\mathbf{x} \rightarrow \Theta)$ while radiance that arrives to this point as $L(\mathbf{x} \leftarrow \Theta)$. The same notation is also used to denote radiance transport between two different points $\mathbf{x}, \mathbf{y}$ in space ($L(\mathbf{x} \rightarrow \mathbf{y})$).

The appearance of an object varies depending on its material and the angle the light illuminates it. The reflectance properties of a material can be captured using Bi-directional Reflectance Distribution Functions (BRDFs). For a point $\mathbf{x}$,

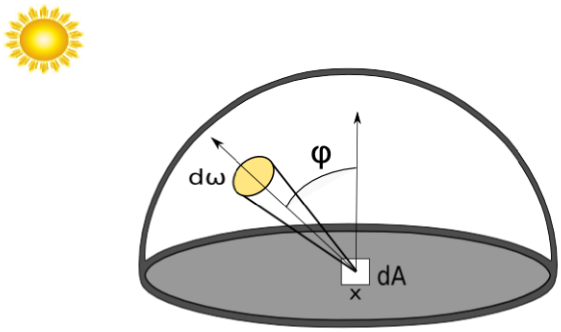


Figure 5.1: Radiance $L(\mathbf{x}, \Theta)$ measured at a point $\mathbf{x}$ in the 3D domain.

the BRDF function can be defined as:

$$f_r(\mathbf{x}, P \rightarrow \Theta) = \frac{\mathrm{d}L(\mathbf{x} \rightarrow \Theta)}{L(\mathbf{x} \leftarrow P)\cos(N_{\mathbf{x}}, P)\mathrm{d}\omega_P}, \tag{5.3}$$

where the directions P and Θ are the incoming and exitant light directions respectively. The term $\cos(N_{\mathbf{x}}, P)$ is the cosine of the angle between the incident light direction P and the vector normal to the point $\mathbf{x}$. Light transport algorithms that make use of BRDFs assume that light arrives and leaves off the same point on a geometrical object. In reality, light might scatter on the surface or get transmitted in the object if it is transparent (e.g. glass). These more complex features can be modelled using Bi-directional Scattering Distribution Functions (BSDFs) [BDW80].

5.2.2 Physically based rendering

Global illumination methods is a class of algorithms used for simulating the transport of light energy in a VE which is composed of surfaces of different scale and properties. At every point in the environment, all possible incoming directions contribute for computing a radiance value at this point. Specifically, for any point $\mathbf{x}$, the radiance that is leaving this point, $L(\mathbf{x} \rightarrow \Theta)$, is equal to the self-emitted radiance ($L_e(\mathbf{x} \rightarrow \Theta)$) (e.g. $\mathbf{x}$ may lie on a light source) and the radiance that is reflected to $\mathbf{x}$ from all incoming contributions ($L_r(\mathbf{x} \rightarrow \Theta)$). The preservation of radiance in the VE can be described using Kajiya's equation [Kaj86] as follows:

$$L(\mathbf{x} \rightarrow \Theta) = L_e(\mathbf{x} \rightarrow \Theta) + \underbrace{\int_{\Omega_x} f_r(\mathbf{x}, P \rightarrow \Theta) L(\mathbf{x} \leftarrow P)\cos(N_x, P)\mathrm{d}\omega_P}_{L_r(\mathbf{x} \rightarrow \Theta)} \tag{5.4}$$

where the integration is done over the hemisphere Ω_x that surrounds point $\mathbf{x}$ and $f_r(\mathbf{x}, P \rightarrow \Theta)$ is the BRDF function at that point. An alternative way to express the outgoing radiance $L(\mathbf{x} \rightarrow \Theta)$ is to consider incoming radiance contributions to point $\mathbf{x}$ from all visible objects in the scene. This formulation can be obtained using the area differential measure (see formula A.2) and integrate equation 5.4.

It is usual in global illumination algorithms, the reflected radiance $L_r(\mathbf{x} \rightarrow \Theta)$ to be written as a direct and an indirect illumination contribution. Direct illumination considers light that arrives directly from the light source while the indirect illumination part accounts for light that reaches point $\mathbf{x}$ after interacting at least once with other surfaces in the scene (see also Figure 5.3).

5.2.3 Monte carlo techniques

Numerical approximation of integrals is very important in a wide range of applications. For example, estimating probabilities or moments of random variables cannot, in most cases, be computed analytically, thus a numerical estimation is required. Monte Carlo methods are traditionally used for estimating multi-dimensional integrals where non-stochastic numerical methods give higher numerical error. The error of non-stochastic methods is of the order of $\mathcal{O}(n^{-1/d})$ where d is the dimension of the space and n is the number of discrete points needed for the estimation. For a function f defined in an interval $[a, b]$ the integral:

$$I = \int_a^b f(x)dx, \tag{5.5}$$

can be approximated using N sample points $x_1, x_2, \cdots, x_N$ from a probability density function (PDF) g. The estimator of this integral takes the form:

$$\langle I_N \rangle = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{g(x_i)}. \tag{5.6}$$

The estimator $\langle I_N \rangle$ is a random variable whose expected value $E_g(\langle I_N \rangle)$ equals the integral I. At the limit $n \rightarrow \infty$, the weak law of Large numbers states that $\mathbb{P}\big(|I - \langle I_N \rangle| > \epsilon\big) \overset{N \rightarrow \infty}{\longrightarrow} 0$ for every $\epsilon > 0$. In other words, for sufficiently large sample size the estimator 5.6 gives more accurate approximations of the unknown integral I. The error of the estimator can be quantified by considering its variance:

$$V_g(\langle I_N \rangle) = \frac{1}{N} \int_a^b \big(\frac{f(x)}{g(x)} - I\big)^2 g(x)\, dx. \tag{5.7}$$

Of course, in the above formula, the integral I is not known thus the variance of the estimator cannot be computed analytically but it can be estimated using an estimator of the same form as the one in formula 5.6. The process of obtaining a sample of N points from a PDF g allows the estimation of both the integral and the variance of the estimator. Formula 5.7 shows that as the sample size increases the standard deviation of the estimator decreases proportionally to $\frac{1}{\sqrt{N}}$. This means that the error in Monte Carlo integration could be halved if the sample size can be increased four times the original sample size N. A variance reduction can also be achieved if someone selects the PDF g carefully. For example, a frequently used methodology is to take samples $g(x_j) \propto f(x_j)$, a technique known as *Importance sampling*.

5.2.4 Light transport simulation methods

This section presents the basic strategies for simulating light transport. For a detailed discussion of the material presented here the interested reader is referred to Phar and Humphreys [PH10] and Dutré *et al.* [DBB06]. These references also provide an in depth discussion of alternative algorithms and techniques that are used in global illumination problems.

Classic ray tracing

Ray tracing techniques are popular methods for rendering physically-based images. These methods model the interaction of light with the surfaces of an environment using infinite straight lines or *rays* that begin from a reference view point. Rays can be compactly parametrised as $\mathbf{p}(t) = \mathbf{o} + t\overrightarrow{\mathbf{d}}$ where, $\mathbf{o}$, $\overrightarrow{\mathbf{d}}$ are the origin point and direction vector of the ray respectively. Rays pass through a point $\mathbf{p}$ that is sampled on an image plane and intersect with the surface which gives the minimum $t \in [\epsilon, \infty]$, where $\epsilon > 0$ is a positive threshold. The ϵ value is a small number that avoids early hits of the ray with the surface it originates and is used to counteract the effect of floating point numerical errors. The radiance value at the point $\mathbf{p}$ equals to the radiance $L(\mathbf{x} \rightarrow \overrightarrow{\mathbf{xp}})$ that leaves point $\mathbf{x}$ towards $\mathbf{p}$ (see Figure 5.2). The first papers [App68, Whi80] that used this technique accounted only for computing direct illumination (i.e. tracing rays from $\mathbf{x}$ to the light source or *Shadow rays*) while they considered indirect illumination effects (reflections, refractions) by sampling only a simple direction from the hemisphere that surrounds the point of interest. For example, in Whitted's original paper [Whi80] only the specular direction was sampled if the material was specular or the refraction direction when the material

is transparent.

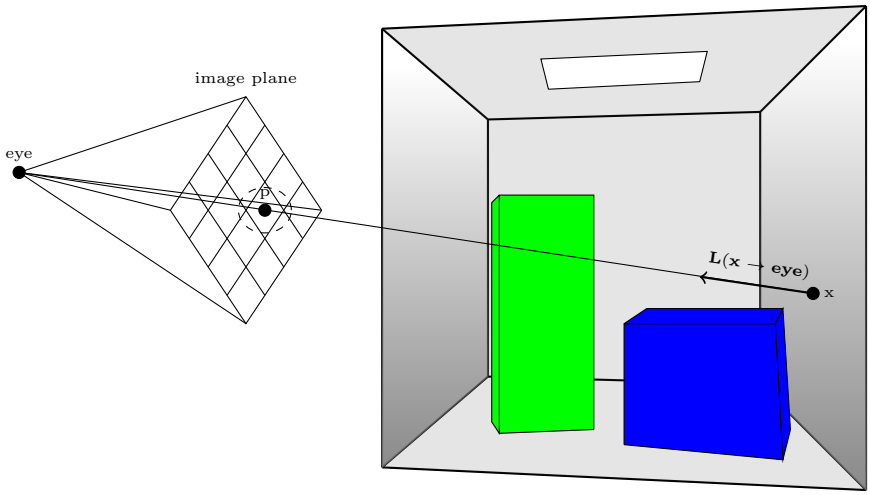


Figure 5.2: Classic Ray tracing.

Distributed ray tracing

Distributed ray tracing methods consider the event where more reflection rays start from an intersection point $\mathbf{x}$ instead of just one as was the case with the classic ray tracing procedures. This methodology gives a much better approximation of the indirect illumination component in a scene. Specifically, by sampling incoming directions P_i using a PDF $g(P_i)$ at the hemisphere surrounding the point $\mathbf{x}$, the reflected radiance $L_r(\mathbf{x} \rightarrow \Theta)$ can be estimated using an estimator of the form given in equation 5.6 by substituting the function f with the integrant quantity at equation 5.4. In that case the number of samples N at the estimator is the number of directions that are sampled at the hemisphere. As far as the the direct illumination is concerned, more shadow rays are traced towards the light sources giving a smoother swadow appearance. Typical examples of this extended ray tracing methodology are given in [CPC84, Coo86]. Effects such as glossy reflections and refractions as well as depth of field and motion blur can be implemented using this technique.

Path tracing

Path tracing is a methodology first introduced by Kajiya as a numerical solution to the rendering equation that was presented at the same article [Kaj86]. For any point $\mathbf{x}$ in the environment, the incoming radiance $L(\mathbf{x} \leftarrow P)$ that appears at the

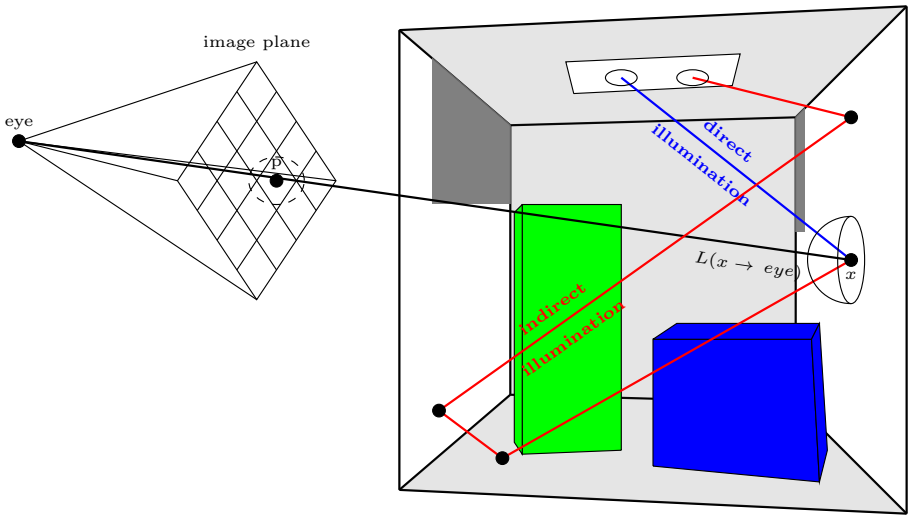


Figure 5.3: Path tracing.

integrant of equation 5.4 can be considered as the radiance that starts at another visible point $\mathbf{y}$ in the scene towards the direction $-P$. In other words, if the points $\mathbf{x}$, $\mathbf{y}$ are visible to each other then $L(\mathbf{x} \leftarrow P) = L(\mathbf{y} \rightarrow -P)$. The radiance computation at $\mathbf{x}$ is therefore dependent on the radiance computation at point $\mathbf{y}$ and this procedure results to a recursive path of rays traced in the scene and contributing to the computation of radiance at the path's initial interesection point. This recursive procedure is stochastically terminated using a process known as *Russian Roulette* [KA88]. At the beginning of the light transport simulation a path termination probability is defined as $\mathbb{P}_{\mathrm{T}}$. This probability is compared to a sampled floating point number $r \in [0, 1]$ at every interesection point. If $\mathbb{P}_{\mathrm{T}} > r$ the path is terminated and the radiance at the last interesection point is accounted for the computation of the radiance at the previous recursion level (the previously intersected point). This procedure is depicted in Figure 5.3. If the probability $\mathbb{P}_{\mathrm{T}} < r$ the path tracing continues to the next interesection point. Another option is to terminate paths after a fixed number of intersections but this methodology introduces bias that is visible through artefacts present at the rendered image. Path tracing provides all possible indirect illumination effects such as interefelctions between surfaces, colour bleeding, caustics, etc. It is widely applied technique in industry and academia and provides a complete solution to the rendering equation (see formulation 5.4).

5.3 Preparation of visual stimuli

This section describes the preparation of the visual stimuli for the implementation of the experiments of this thesis. The description starts with the selection of an image quality metric and a generic cost function that is required for estimating the theoretical cost for better quality visual cues. Finally, a methodology based on the VDP computational model is described for discarding quality levels that elicit no perceptual visual difference to the viewer.

5.3.1 Quality metric for visual cues

For visuals, resolution was chosen as the variable that modifies quality. This was preferred over other metrics of quality because it is the most straightforward while the visual differences are distinct and easy to understand. A number of standards already exist for resolution and it is possible to be abstracted from the underlying algorithm used for image synthesis. Furthermore, the computational cost of rendering an image is, for many rendering algorithms, a linear function of the image resolution.

All the images were computed using path tracing [Kaj86] due to the accuracy and straightforward nature of the computation. The software was implemented from the ground up and no available rendering software was used. The CAD[1] models for the experimental studies of this thesis were designed using Autodesk Maya while photos from real scenes were captured author used for designing the models. The machine used for rendering these images was an Intel Xeon E7450 at 2.40 GHz with a total of 24 physical cores and 64 Gigabytes of RAM. Figure 7.1 depicts all the rendered images. Both the models and the rendered images were made specifically for the experimental studies of this thesis.

The objective was to create realistic scenes that present physically correct illumination and material properties, and may be representative of future rendering systems. All the images were rendered to convergence to generalise this work to any possible algorithm. The experimentation framework included the computation of 240 images which varied in resolution from 16×9, to the highest resolution at Ultra High Definition (UHD) or 3840×2160 using a fixed aspect ratio of $16 : 9$ for all the images (i.e. 16×9, 32×18, $\cdots$, 3840×2160). Figure 5.4 depicts the process of generating all the images. The lowest resolution was chosen to reflect the level at which humans find it difficult to identify images [Tor09].

In order not to introduce any bias due to the presented size, all the different

[1] Computer Aided Design.

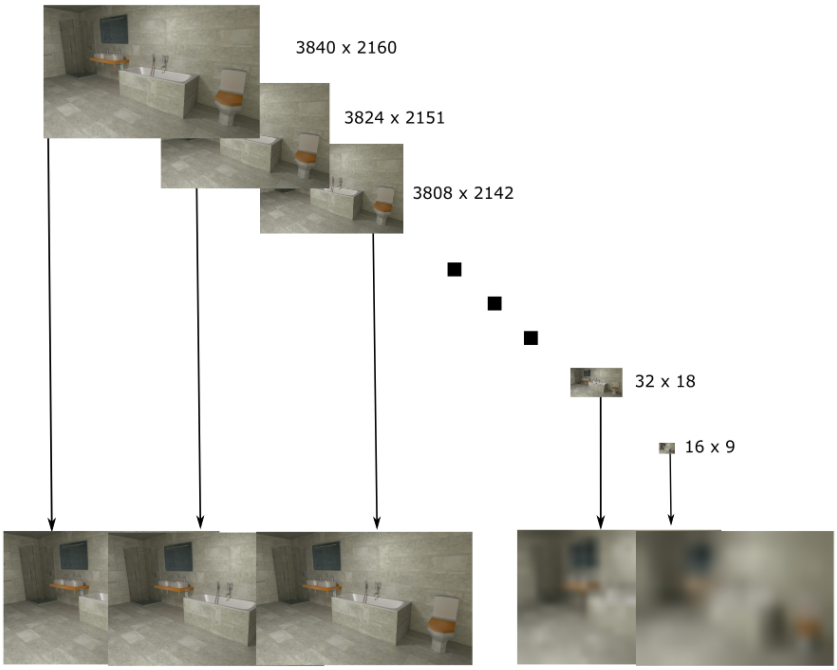


Figure 5.4: Images rendered at different resolution and scaled to the highest resolution displayable by the visual display (3840×2160).

quality images were resized to the same resolution, namely, UHD resolution used by the display hardware. This resizing process was implemented using a bi-cubic interpolation kernel while anti-aliasing and colour dithering filters were used in order to keep the quantisation error to a minimum. Bi-cubic interpolation was preferred over other image scaling methods because it produces smooth images and its application has no significant computational cost relative to the overall rendering costs.

5.3.2 Cost estimation for visual cues

The computation time needed for obtaining an image of the sequence of 240 images can be generally estimated as follows:

$$C_k^V \approx P_k \cdot L, \qquad k = 1, 2, \cdots, 240, \tag{5.8}$$

where $P_k = 16 \cdot 9 \cdot k^2$ is the number of pixels of the $k-$th image and $L = C_{240}^V / P_{240}$ is the time needed for the computation of an individual pixel. It is worth mentioning that this estimation, strongly depends on the available hardware, the algorithm used and the optimisations applied at an algorithmic and software level. Other parameters that affect the computation time include, but are not limited to, the scene

complexity, material and texture properties. However, in order for the results of the experimental studies ***E1*** and ***E3*** proposed in the methodology to be generalised, the assumption that the computational cost is varying linearly with resolution is made.

Moreover, the problem is decoupled from real time measurement and is considered in terms of normalised cost that can be given by equation 5.8 using the normalisation factor $\frac{1}{C^V_{240}}$. This results in visual levels with costs independent of the underlying algorithm used for the computation. Therefore, in what follows, the term *visual cost* is used to denote the quanity given by:

$$C^{N_V}_k = \Big(\frac{k}{240}\Big)^2, \quad k = 1, 2, \cdots, 240, \tag{5.9}$$

This formula yields no real cost but a theoretical cost that is used to assess the different computional requirements of the 240 visual quality levels in a consistent way independent on the underlying algorithm used for rendering the visual cues.

5.3.3 Visual selection criteria

The HVS's propensity to observe visual differences at lower resolutions has led to consider only those visual stimuli that can be considered perceptually distinguishable. For the correct implementation of the experimental procedure, only the subset of images that elicit visual differences at perceptual level were needed. Images that elicit the same perceptual response and have different costs might lead to false conclusions in the resource allocation framework. In order to obtain a set of perceptually distinguishable images a visual perceptual metric was used.

The latest version of the HDR-VDP computational model (HDR-VDP-2.2) [NMDSLC15] was used for percetually assessing the rendered images. This model is a widely used objective metric for detecting perceptual differences between High Dynamic Range (HDR) or Low Dynamic Range (LDR) image pairs. The model provides the Q correlation measure, a numeric score that ranges between $0 - 100$. Low Q scores indicate apparent visible differences between the input images while two images with high Q value are considered perceptually indistinguishable. For the selection of distinguishable visual stimuli, pairwise comparisons between the highest resolution image (UHD) and the 239 other rescaled images were performed using the LDR mode of the metric. The averaged Q scores for all the six scenarios used in experiment ***E1*** are depicted in Figure 5.5. As expected, the results follow a logarithmic trend indicating that at higher levels participants struggle to find apparent differences.

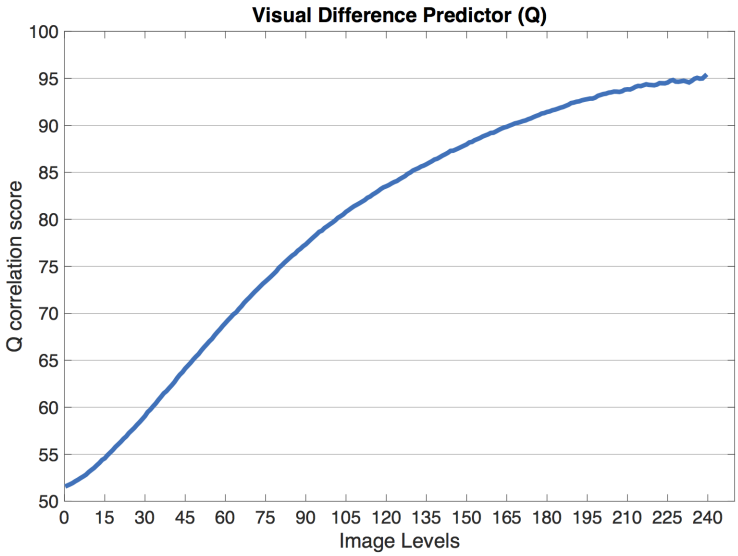


Figure 5.5: Average values of the Q correlation results obtained from the VDP model for all six scenarios that were considered for the ***E1*** experimental study.

In their study, Varshney and Sun [VS13] show that the internal representation of a stimulus scales in logarithmic fashion for increasing stimulus magnitudes. The authors argue that any range of stimulus physical intensities is mapped through a log-curve to a finite set of perceptual points that consitute distinct sensory levels for the human observer. This mechanism resembles signal quantization where the internal perception points are uniformly spaced and very high intensity values have little or no perceptual impact to the internal representation (upper quantization boundary). This argument is true not only for the visual domain but also for other human sensory systems [VS13].

This fundamental psychophysics result reported by Varshney and Sun was used to discretise the Q values to 80 discrete levels that elicit noticeable perceptual differences to the user. This number is the minimum possible number of levels in order for the resulting set of images to include the majority of the common resolution standards that are frequently used in High Definition Television (HDTV) and Standard Definition Television (SDTV) applications. These are:

- Quarter High Definition (QHD or 960×540)
- High Definition (HD or 1280×720)
- Full High Definition (FHD or 1920×1080)
- Ultra High Definition (UHD or 3840×2160)

5.4 Summary

This chapter explained how the visual stimuli of the experimental studies ***E1*** and ***E3*** were rendered. The computation of these images was based in path tracing, a widely used algorithmic technique for obtaining global illumination features. Cost estimation for visuals is based on formula 5.9, which for higher image resolution levels (k) yields a higher theoretical cost $C_k^{N_V}$. Finally, visual selection criteria using the VDP model were discussed. These criteria were used to discard those visual stimuli that are perceptually indistinguishable and have different theoretical costs.

Chapter 6

Auditory pipeline and materials

6.1 Introduction

This chapter describes the method used for rendering and delivering audible cues. The sense of hearing is one of the major senses in the experimental framework along with the sense of sight (see chapter 4). Physically based rendering of auditory stimuli is used for the computation of the audio cues needed for the implementation of the experiments ***E1*** and ***E3*** in chapters 7 and 10 respectively. This chapter starts with a brief introduction of the process of auralisation in section 6.2 and its role in VR applications. The preparation of the auditory stimuli for the implementation of the experimental studies is given in section 6.3. This includes the selection of a metric for assessing the quality of the audio tracks and the definition of a cost function for assigning higher computational cost depending on the quality of the stimulus. As was the case with the rendered images in the previous chapter, perceptually similar audio tracks are not included in the final set of auditory stimuli used for the experiments.

6.2 Acoustics rendering background

This section presents the basic concepts needed for modelling sound transport in a VE. Similar to the process of visualisation which is used to describe the rendering and delivery of visuals in the form of images or videos, the process of *Auralisation* involves the simulation of sound waves in a domain of interest and the delivery of an audible sound stimulus to the user. The rendering of audio stimuli requires a set of prerequisities before the actual simulation stage. These include: a well defined geometry that represents the VE where the sound propagates within, material ab-

sorption properties under a range of frequency bands, source and listener positions in the domain of interest and finally an anechoic sound signal that will be filtered appropriately to contain all the reverberation effects due to the geometrical features of the VE. The process of auralisation results to audio as of being listened in the real environment.

The output of every sound transport rendering algorithm is a Room Impulse Response (RIR), a digital signal that encodes delays and attenuations of the sound waves as they interact with the surfaces of the environment before approaching listener's position. The RIR is used to enrich an anechoic audio stream with reverberation effects through convolution with the anechoic stream. Appendix B defines the convolution of signals and explains its basic properties in the context of signal processsing. In the next, a discussion about the properties of a RIR is given followed by a brief introduction of the simulation methods used for modelling sound propagation.

6.2.1 Room impulse response

A room impulse response is the output of a Linear and Time Invariant (LTI) filter when a Dirac function $\delta(x)$ is given as input. A brief overview of LTI filters is also given in Appendix B. The Dirac delta function is defined as a function that gets the zero value everywhere except when its argument is zero where it gets the infinity value. The properties of such a function are given as follows:

$$\begin{aligned} \delta(x) &= 0 \text{ when, } x \neq 0 \\ \int_{-\infty}^{\infty} \delta(x)dx &= 1 \\ \int_{a}^{b} \delta(x-z)f(x)dx &= f(z) \text{ when, } a \leq z \leq b. \end{aligned} \tag{6.1}$$

In the context of auralisation the room plays the role of the LTI filter and the output impulse response captures the interactions of the sound waves with the surfaces that compose the virtual scene. The output impulse response is convoluted with an anechoic sound (*dry* sound) and produces an audio stream (*wet* sound) that contains reverberation effects and attenuations [Kut09]. It is important to note that for different positions of the listener or the sound emitting source a new RIR should be modelled to encode the interactions of the sound wave with the geometry between the two new positions. The waveform representation of an output RIR is depicted in Figure 6.1.

The RIR signal is composed of three distinct regions that have different acous-

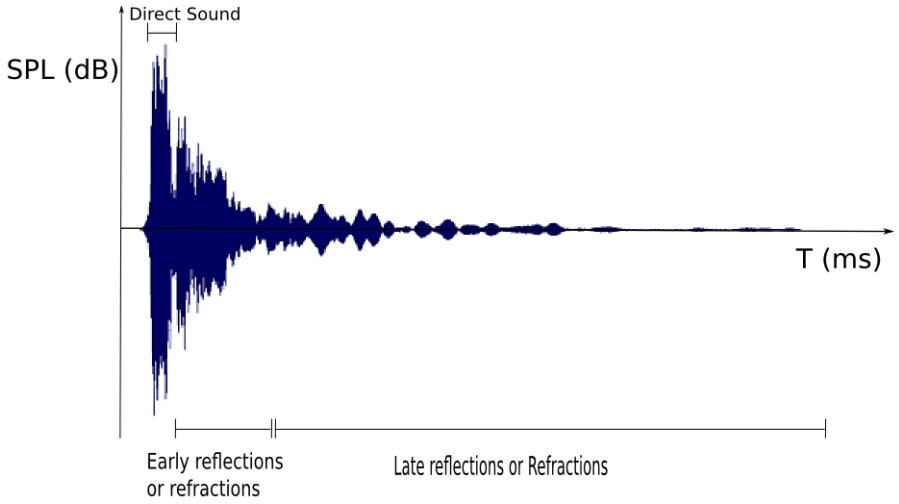


Figure 6.1: A simulated Room Impulse Response. The y-axis represents sound pressure level in dB while the x-axis represents time measured in milliseconds.

tic properties and duration. The first part comprises the direct sound paths that reach listener's position without interacting with any of the surrounding surfaces. This part of the RIR is very short and contains the maximum instantaneous acoustic energy that is received by the listener. The second part, known as early reflections and refractions, contains acoustic paths that range between 5 to 100 ms and contains information about the directionality and the pressure levels of the emitted sound wave [Dur94, GA97]. This part allows the listener to perceive the scale and the reflective properties of the room that the sound propagates within [BH98, Nie92]. Finally, the last section, known as late reflections and refractions, captures all these acoustic paths that reflect and diffract multiple times before reaching the listener. The sound pressure level significantly decreases in the last part thus it is more difficult to get perceived by the listeners with the exception of places where reverberation effects hold for some time before sound fades to an inaudible level (e.g. cathedrals, caves, train tunnels, etc.).

Sound propagation rendering techniques focus on accurately modelling the early relections and refractions part as it strongly affects users' listening experience [WR93]. Less accurate approximation methods are followed for the late reflections and refractions [SHHT86].

6.2.2 Sound transport simulation methods

A sound transport algorithm is used to simulate the various interactions of the sound waves with the environment. These interactions include reflections, refractions and diffractions upon the surfaces and objects that compose the VE. A common methodology for simulating sound wave propagation phenomena is based on ray tracing methods.

As was the case with visuals, wave propagation can be simulated using infinitely thin rays. Siltanen *et al.* [SLKS07] investigated a model that describes sound energy transport in a VE in the form of a rendering equation similar to the one introduced by Kajiya for light propagation [Kaj86]. This equation can be written as follows:

$$L(\mathbf{x}' \to \Omega) = L_e(\mathbf{x}' \to \Omega) + \int_G R(\mathbf{x}, \mathbf{x}', \Omega) L(\mathbf{x}' \leftarrow \overrightarrow{\mathbf{xx}'}). \tag{6.2}$$

The term $L(\mathbf{x}' \to \Omega)$ denotes the outgoing acoustic energy from point $\mathbf{x}'$ in the direction Ω. If the point $\mathbf{x}'$ lies on a sound source the term $L_e(\mathbf{x}' \to \Omega)$ is non zero and is used to denote the acoustic energy that originates from that point otherwise it takes 0 value. The integration is done over all surface points $G \subset \mathbb{R}^3$ that compose the geometry. The vector $\overrightarrow{\mathbf{xx}'} = -\overrightarrow{\mathbf{x}'\mathbf{x}}$ denotes the unit vector $\frac{\mathbf{x}-\mathbf{x}'}{\|\mathbf{x}-\mathbf{x}'\|}$. Figure 6.2 gives a geometric representation of the terms presented in equation 6.2. The term:

$$R(\mathbf{x}, \mathbf{x}', \Omega) = V(\mathbf{x}, \mathbf{x}') f_r(\mathbf{x}', \overrightarrow{\mathbf{xx}'} \to \Omega) G(\mathbf{x}, \mathbf{x}'), \tag{6.3}$$

includes the BRDF term at the point $\mathbf{x}'$, denoted as $f_r(\mathbf{x}', \overrightarrow{\mathbf{xx}'} \to \Omega)$ while the term $V(\mathbf{x}, \mathbf{x}')$ is an indicator function that shows if the points $\mathbf{x}, \mathbf{x}'$ are visible to each other. It can be written as:

$$V(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & \text{if, } \mathbf{x} \text{ and } \mathbf{x}' \text{ are mutually visible.} \\ 0 & \text{if, } \mathbf{x} \text{ and } \mathbf{x}' \text{ are not mutually visible.} \end{cases} \tag{6.4}$$

$G(\mathbf{x}, \mathbf{x}')$ describes the effect of the geometry on the acoustic energy that flows from point $\mathbf{x}$ to point $\mathbf{x}'$. This term can be written as:

$$G(\mathbf{x}, \mathbf{x}') = \underbrace{(N_{\mathbf{x}} \cdot \overrightarrow{\mathbf{xx}'})}_{\cos\Theta_1} \underbrace{(N_{\mathbf{x}'} \cdot \overrightarrow{\mathbf{x}'\mathbf{x}})}_{\cos\Theta_2} \frac{S_{\|\mathbf{x}-\mathbf{x}'\|}}{\|\mathbf{x}-\mathbf{x}'\|^2}, \tag{6.5}$$

where the operator $S_{\|\mathbf{x}-\mathbf{x}'\|}$ represents sound propagation effects over distance $\|\mathbf{x}-\mathbf{x}'\|$. The terms $(N_{\mathbf{x}} \cdot \overrightarrow{\mathbf{xx}'})$ and $(N_{\mathbf{x}'} \cdot \overrightarrow{\mathbf{x}'\mathbf{x}})$ are the cosine of the angles Θ_1 and Θ_2 respectively. These are also depicted in Figure 6.2. It should be noted that equation 6.2 represents the area formulation of the rendering equation. An identical expression

can also be written for the light rendering equation 5.4 by replacing acoustic energy with radiance.

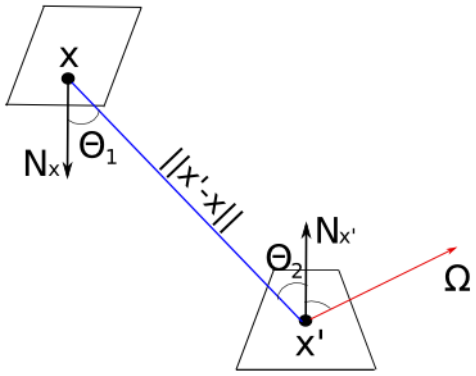


Figure 6.2: Graphical representation of the terms term given by the equation 6.5.

Equation 6.2 describes mathematically the process of tracing rays from a sound source and collecting them to a probe location. The traced rays carry acoustic energy that is propagated in the environment. The rays collected in the probe location have lower acoustic energy after the interactions with the objects of the environment. The location of the virtual probe represents the position of the listener in the VE. In the following, basic techniques for sound transport simulation are proposed. For a complete reference of existing sound spatialisation methods the interested reader is referred to the paper [HHD+12].

Ray-based method

Ray-based methods [KSS68] simulate sound reverberation paths via tracing rays that begin from a sound source and end up at probe's location forming finite paths in the scene (see Figure 6.3). Sound intensity is computed at ray hit points and every ray accounts for a small portion of acoustic energy that is emitted by the source and collected by the receiver after ray's interaction with VE's geometrical features. The success of this method depends on the number of distinct ray directions that can be exploited by a Monte Carlo estimator at every hit point. Sound artefacts may be heard in the resulting impulse response due to under-sampled areas in the simulation domain [L.93]. Aliasing problems can be partially solved by tracing more rays in order to decrease the variance of the estimator and provide an acoustic result with reduced noise.

Ray-based methods can be used to simulate effects such as specular and

diffuse reflections and refractions while simulating diffraction effects is more difficult due to the nature of the ray (thin lines) and can also lead to noise artefacts [LB92]. The advantage of this method is its implementation simplicity and scalability when the number of computing resources increases.

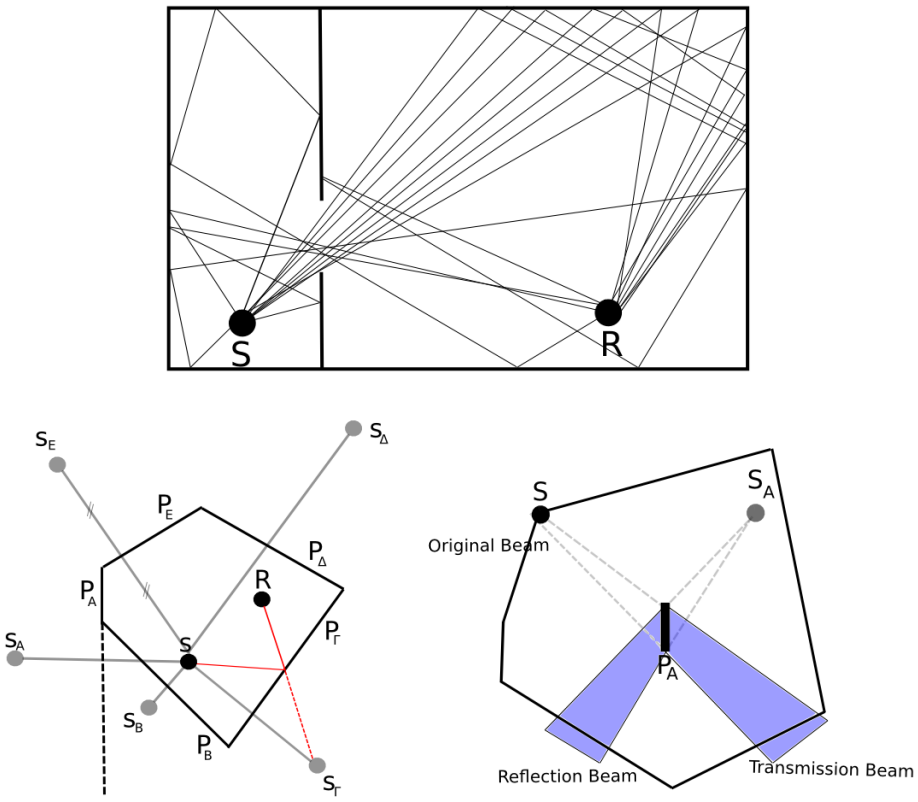


Figure 6.3: **Top:** Simulated paths between the source and the receiver using the Ray-based method [KSS68]. **Left-Bottom**: The generation of new sound sources at the Image source method [AB79]. **Right-Bottom:** Beam tracing method [FTC+04].

Image source method

The image source method as was proposed by Allen and Berkley [AB79] is a sound transport simulation method that computes only the specular reflections of the sound in the environment. The method begins by constructing virtual "images" of the sound source at reflected positions across every surface of the environment (see Figure 6.3). Specular reflection paths can be realised by intersecting a line segment from the source position, S to every reflection plane in the VE. This method recursively gives reflection paths of arbitrary order but the computational complexity

increases exponentially. Specifically, for m surface planes and r reflection implementations, the algorithm complexity increases to $\mathcal{O}(m^r)$. This method is practical when applied to hexagonal domains where a relatively small number of virtual sources is required for obtaining an accurate RIR.

Extensions of the method propose its application to more complex domains using visibility tests for eliminating virtual sources which have no significant contribution for the given listener position [Bor84]. Other studies have coupled the image source method with other sound transport simulation methods. These hybrid methods exploit the advantages of both methods with reduced computational cost [SHTV99].

Beam tracing method

Beam tracing methods simulate acoustic paths in the domain using bundles of rays (*beams*) instead of rays. A beam is modelled as a 2D region containing a number of rays that point at different directions. Beam tracing methods have also been applied to global illumination problems [Wat90], visibility determination [Tel92] and radio wave simulations [For99]. In the context of sound transport modelling, when a beam intersects a surface is clipped and two new beams are constructed. One is a transmission beam that extends the original beam trajectory while the other, called reflection beam, is constructed by symmetrically projecting the transmission beam over the intersected surface (see figure 6.3). This method requires the construction of less virtual sources even for environments with fairly complex geometries while it produces less under-sampling artefacts compared to the ray tracing method as a wider region of directions is covered using a single beam.

On the downside, beam tracing methods require complex geometry calculations particularly in the design of beam surface intersections. This makes the implementation considerably harder as it is difficult to generalise the intersection methodology to various geometries of different scale and shape. Studies have proposed simplification methodologies for overriding geometry complexity by taking into account surface intersections only with a limited number of rays that compose a beam [Lew93].

Funkhouser *et al.* [FTC+04] described a framework based on the beam tracing method that is capable of computing spatialised audio in real time. Their implementation is based on a considerable amount of pre-computation that is intended for firstly, finding and storing adjacency relations of polygons in the input geometry and secondly, discover all the transmission and reflection beam paths.

6.3 Preparation of auditory stimuli

This section describes the preparation of the auditory cues used in this thesis. Similar to the visual cues, a metric was adopted for assessing the quality of the auditory stimuli. This metric is then related to a cost function that assigns theoretical costs to the audio tracks. Better quality cues are assigned higher cost while lower quality cues worth lower cost for selecting them in the resource allocation task.

6.3.1 Quality metric for auditory cues

For the auditory part of the experimental studies ***E1*** and ***E3*** a series of RIRs were computed for the same set of scenarios used for generating the visual images. A ray-based approach was adopted for rendering the acoustic stimuli [SLKS07]. The generated RIRs were convoluted with an anechoic stream for obtaining reverberation effects in the final audio stream. The anechoic stream is a recording of the sound without any reverberation effects.

All the anechoic audio files were recorded in an anechoic chamber, a room designed for allowing only the propagation of the direct sound and eliminating possible reverbaration effects of the sound in the recording (see also section 6.2.1). All the anechoic sounds were captured as uncompressed audio files using .WAV format and had different duration times (between 5 to 10 seconds). As far as the intensity of the delivered sound stimuli is concerned, it was constant at 45 dB in all the experimental studies described in this thesis.

The selected quality metric for audio was the sampling rate of the rendered RIR. All the RIRs were rendered in B-format (Ambisonics) at 352800 Hz sampling frequency as the highest temporal resolution for sound. Such high sampling rates are used in Digital extreme Definition (DxD) applications for editing high-resolution audio tracks.

Similarly to visuals, 240 RIRs were computed at 240 different sampling rates for each of the scenarios. Every rendered RIR was interpolated to the maximum sampling rate of 352800 Hz for playback using bandlimited interpolation and specifically a *sinc* interpolation kernel [SE98]. This process, oversamples the RIR signal as the target sampling rate is significantly higher than the Nyquist frequency. The oversamping operation does not affect the quality of the rendered RIR and it is implemented so as to have RIRs which have the same sampling rate as the anechoic stream before convolving the two signals. The anechoic stream was also oversampled using a *sinc* interpolation kernel at 352800 Hz before convolving with the RIR.

After the sampling rate conversion, every RIR is convolved with a two

channel Head Related Impulse Response (HRIR) signal using the CIPIC HRTF database [ADTA01]. The HRIR signals capture all source localisation cues and encode elevation and azimuth effects of the incoming sound signal to a listener's ears (see also section 2.2.4). These cues are affected by human anatomical characteristics (head size, shoulder, pinna shape, etc.) and the content of the HRIR signal is used to capture the effect of this interaction.

6.3.2 Cost estimation for auditory cues

The normalised computation time needed for rendering the k-th RIR of sampling rate $f_k < f_{240}$, where $f_{240} = 352800$ Hz, can be estimated as follows:

$$C^A_{f_i} \approx \frac{f_i}{f_{240}}, \quad \text{where } k = 1, 2, \cdots, 240. \tag{6.6}$$

The normalisation factor $1/C^A_{f_{240}}$ was directly applied in the above equation to indicate that the estimated quantity represents normalised and not physical computational cost. The non-normalised version of equation 6.6 reflects the fact that a RIR with half the number of samples per second needs approximately half computation time to be rendered. Throughout this thesis the term *audio cost* is used to represent the normalised cost given by equation 6.6.

As was the case with visuals (see section 5.3.2), the above formula gives a coarse estimation of the computation time and other factors might affect the estimation, such as, the reflection properties of the environment's surfaces under a multitude of different frequencies, the relative positions of the listener and receiver inside the geometry as well as the technique used to compute the RIR.

The sampling rates of the 240 audio levels were determined using the inverse of the function given in equation 6.6 and applying, as arguments, the values of the visual costs defined in equation 5.9. Specifically, the k-th RIR has a sampling rate given by:

$$f_k(C^{N_V}_k) = \lceil C^{N_V}_k \cdot f_{240} \rceil, \quad k = 1, 2, \cdots, 240. \tag{6.7}$$

where the $\lceil \cdot \rceil$ operator is used to obtain the closest integer sampling rate with the highest value. In practice, the application of the above equation makes the visual and auditory sensory stimuli comparable in terms of their normalised cost; a practical property that allows to consider visual-audio cost interactions without worrying about which particular algorithm was used for computing the stimuli of either sense. Cost interactions between the senses are discussed in detail in section 7.2 of the next

chapter.

The resulting sequence of audio levels included RIRs of progressively higher sampling rate until the last rendered RIR. Lower auditory levels could be heard as "dim" sounds because only low frequencies are present in the respective signals while higher auditory levels were rich in harmonics making the sound perceived brighter and clear. This is analogous to the visual domain where higher quality levels are sharp images while lower quality images are perceived "blurry" due to the absence of high frequency features (edges). The progressive existance of high frequency components makes the process of distinguishing audible differences between higher quality audio pairs quite difficult. As was the case in the visual domain, it was important to retain those audio levels that elicit different perceptual stimulations to the average user.

6.3.3 Auditory selection criteria

To the best of the author's knowledge, there is no analogue of the HDR-VDP in acoustics that can be used to estimate perceptual differences between two audio tracks. There are many parameters that affect the process of finding audible differences, such as: loudness, pitch, duration, room clarity, instantaneous sound energy, content (e.g. music or speech), sound reflection properties of the room, etc. It is also important to note that identical audio stimuli do not necessarily elicit the same perceptual response on a person by person basis [Est76], making the development of such an objective metric more complicated.

However, the frequency JND in Acoustics is 1Hz for complex sounds in the range of 500–1000 Hz while it progressively increases for sounds of higher frequencies [KBM08]. This means that it is easier for a human listener to distinguish two different sounds when their content is at low and mid-frequencies rather than at higher frequencies of the audible spectrum.

For the experimental studies **E1** and **E3**, 80 auditory levels were kept using the same log spacing distribution as was applied in the visual domain using the VDP results. This log-spacing keeps the majority of the low frequency RIRs where perceptual differences are clearly audible while it discards many high frequency RIRs where the audio stimuli start becoming perceptually indistinguishable.

The 80 audio tracks that kept from the original set of 240 convolved sounds have frequency content that satisfies the JND requirement of 1Hz apart for understanding the perceptual differences between them. Therefore, they can be considered perceptually distinguishable to the average user.

6.4 Summary

This chapter explains the process followed for rendering auditory cues in this research. Ray tracing was employed for computing RIRs in the experimental studies **E1** and **E3**. Sampling rate is used as a quality metric for assessing the RIR signals after their computation. Higher sampling rate RIRs are assigned higher theoretical cost $fC^A_{f_i}$. The JND for frequency was used to evaluate the convolved sounds in terms of finding perceptual differences between pairs of audio tracks. The following chapter describes how the visual and auditory materials of this and the previous chapter are used for the implementation of an experimental study for capturing users' resource allocation preferences in bi-modal VEs.

Chapter 7

Resource allocation for bi-modal virtual environments

7.1 Introduction

The work presented in this chapter examines how the budget of available resources affects the perceived fidelity of an audio-visual experience. The objective is to study people's subjective preferences when distributing a limited budget for quality improvements in the displayed visual and auditory stimuli. A model is presented that can be used for distributing resources at generic scenarios. This model is based on the subjective experiment whereby participants' quality preferences when allocating resources were captured and analysed. This prediction model is then evaluated with different, previously untested budgets and previously tested and untested scenarios. Rendering systems could make use of such a model to direct resource allocation in bi-modal VEs. Given a computational budget, resource balancing estimations between audio-visual qualities can be computed on the fly.

The rendering methodology and the techniques used for evaluating the cost of the visual and auditory stimuli were described in chapters 5 and 6 of this thesis. The rendered images and audio tracks represent quality levels in each of the two senses. Participants can alter the quality in each of the two senses using the controls of a GUI where improvements in vision and/or audio correspond to increased computational cost allocated from the total budget. The assigned budget corresponds directly to quality improvements. In turn, these quality improvements in vision and audio directly correspond to higher computational cost that is allocated from the available budget.

The experimental procedure requires the users to move interlinked sliders

that modify the quality of the displayed graphics and acoustics interactively. When one is increased the other is automatically decreased. The experiment was conducted with five distinct budgets across five different scenarios. The experimental design is within participants and each participant is asked to choose the best perceived quality for all budgets and scenarios. The presentation of all the 25 possibilities was randomised to avoid any potential ordering bias on behalf of the participants during the experimental session.

The content of this chapter is organised as follows. Firstly, the terms "cost" and "budget" are defined and explained in section 7.2. Secondly, the experimental layout that includes the design, materials, participants and procedure is given in section 7.3. Thirdly, results and their statistical analysis is given in section 7.4. Fourthly, two statistical models are proposed and validated in a subsequent experimental study, this study is described in section 7.5. Lastly, section 7.7 gives a detailed discussion of the current limitations of this work and possible ways to overcome these limitations.

7.2 Visual-auditory cost interactions

For the implementation of this experimental study, the normalised cost functions were used for both visual and auditory levels. These cost functions are given by formulae 5.9 and 6.6 for visual and auditory stimuli respectively. Moreover, the two senses follow a similar cost distribution as the costs for visual stimuli were used to find the sampling frequencies of the 240 original auditory levels. These assumptions overcome the problem of the unbalanced physical computation time needed for rendering a high quality image compared to that of the high quality RIR in a VE. These assumptions do not affect the generic scope of the study which is to find the relative importance of each of the two senses in terms of the percentage of the total budget that is devoted to each of them.

The notion of "budget" as it is used throughout this thesis adapts to the idea of the normalised cost and represents a theoretical quantity that is distributed amongst the costs of a visual and an auditory quality level. Five different theoretical budgets are used; these are shown in Table 7.1 along with their notation letters. The number of different budget sizes allows investigation on how the participants' allocation preferences scale for very small budget sizes (B_1) where quality improvements are constrained, up to larger budget sizes (B_4, B_5) where the user is able to select relatively high quality levels in both senses.

The budget selection was determined such that the maximum quality level

Budget letter	$\mathbf{B_1}$	$\mathbf{B_2}$	$\mathbf{B_3}$	$\mathbf{B_4}$	$\mathbf{B_5}$
Budget	0.0625	0.11	0.25	1	1.12
Total Number of Levels	28	38	48	80	48
Vision (resolution)	QHD	HD	FHD	UHD	UHD
Audio (kHz)	22.05	38.8	88.2	352.8	352.8

Table 7.1: All the theoretical budgets used in ***E1*** experimental study. The first row shows the budget notation letter. The second row shows the value of the theoretical budget. The third row gives the number of levels (out of 80) that remain when a budget is applied. The fourth row gives the resolution standard that is obtained when the maximum level is selected in vision (see also section 5.3.3 for more details). The last row gives the sampling frequency standard that is obtained when the maximum level is selected in audio.

in vision or audio corresponds to one of the commonly used resolution and sampling rate technology standards. These standards are given in the fourth and fifth line of the table 7.1. The shorthand notations QHD, HD, FHD and UHD are acronyms for quarter high definition, high definition, full high definition and ultra high definition respectively and their values are given in section 5.3.3. Table 7.1 also shows the total number of quality levels that remain when one of the five budgets is applied in the experiment.

Budgets B_1, B_2, B_3 use a subset of the original set of quality levels while budget B_4 contains the original number of perceptually distinct levels which is 80. Budget B_5 does not add new quality levels but instead the user starts improving the quality from medium visual-auditory quality to compensate for the extra budget amount, $B_5 - B_4$, that is given. B_5 is the only budget size that allows the user to select the maximum quality in either of the two senses and at the same time receive a medium to high quality stimulus from the other sense.

As can be seen from table 7.1 budget B_5 has less number of levels than B_4 because low quality levels are not allowed to be obtained. The use of budget size B_5 was the only way to increase the budget size more than B_4 without using images of resolution more that UHD which is the maximum resolution that could be delivered by our available display (4K monitor).

7.3 Experimental layout

This section gives details about the experimental layout of this study. Specifically, the design, hardware and software materials used, information about the partici-

pants and finally the procedure that was followed during an experimental session are thoroughly discussed.

7.3.1 Design

The experimental design is organised such that participants can adjust the quality of a rendered image and an audio track using the controls of a user interface. Participant's preferences were recorded when different budget sizes are available for allocation. The design of this experiment includes the effect of two independent variables, namely, *scenario* and *budget* on two dependent variables. The two dependent variables are the percentage of the budget that is devoted to the graphics and acoustics quality respectively. It is clear that the value of one of the dependent variables completely adjusts the value of the other (negatively correlated).

The independent variable *budget* has five different levels (budget sizes). These are given in table 7.1. The independent variable *scenario* has five different levels (scenarios). These scenarios are given in Figure 7.1 and are discussed in the following. One scenario, the bathroom, is used for training but not for data collection, therefore, it does not count as a level of the independent variable *scenario*.

7.3.2 Materials

For this study a 28″ Samsung U28D590 ultra HD LED monitor was used to display the LDR images at resolution 3840×2160 while a Dell UltraSharp 2007WFP 19″ LCD monitor was used to display the GUI. All the sounds were delivered binaurally using a set of Sennheizer HD 380 PRO headphones.

The distance from the participant's head to the main display and the rotation angle needed to watch the contents of the secondary display are shown pictorially in Figure 7.2. These follow the guidelines of the ITU-R BT.500-13 standard [ITU12] for adjusting viewing conditions in subjective evaluations using HDTV and SDTV display panels. The experimentation procedure was conducted under dark ambient luminance conditions in a silent room.

A total of six scenarios were used for the experiment ***E1***. These are labeled: Bathroom, Car, Kitchen, Kitti, Restaurant and Yard. The Bathroom scene was used only for training before the formal experimental session commenced and it was not used for collecting experimental data during this experiment. Images of the six scenarios are given in Figure 7.1.

The sound stimulus in the Kitchen scenario was assumed to originate from a kettle that was visible to the subjects. The Yard scenario consisted of the sound of

Figure 7.1: All the visual images of the scenarios used in ***E1*** experimental study. From **Left** to **Right** and **Top** to **Bottom**: Bathroom, Car, Kitchen, Kitti, Restaurant, Yard.

the engine ignition of a visible lawnmower at the right hand side of the screen. The Kitti sound stimulus was a hymn where the sound source originated from behind the main church tableux. For the Restaurant scenario the sound of an employee's voice delivering food behind the wall was used while there was clatter noise coming from the same position. The Car scenario sound includes both engine ignition and the putting on of a seat belt. The Bathroom scenario included the sound of water flowing in the shower. The objective was to present both indoors and outdoors scenes where the audio stimulus varied from mechanical everyday sound to human speech and melody.

A custom GUI application was developed for the subjective study ***E1***. The interface included two main windows, one displayed the images in full screen at the main monitor and the other contained a window with the basic controls for adjusting

Figure 7.2: **Top:** Hardware set-up of the ***E1*** experimental study. The viewing distance to the main monitor and the angle needed to see the contents of the secondary monitor are given in [ITU12]. **Bottom:** Photo taken during preliminary tests.

the graphics and acoustics quality. The interface design was based on a pilot study with six participants. Their feedback assisted to improve the GUI before starting the experimental study. The participants of the pilot study did not take part in the experimental study for avoiding bias on the results. Two slider thumbs were used to dynamically change the quality of the presented visual and auditory stimuli by taking values that correspond to different quality levels.

The effect of different budget sizes is shown in the GUI by colouring red the portions of the slider bar which contain quality levels that cannot be presented with the given budget. An example of the GUI configuration is given in Figure 7.3. This image presents an instance of the medium budget B_3 and an instance of the smallest budget size B_1 used while it depicts all the controls that composed the GUI. The visual slider bar, along with its label, were randomly interchanged with the auditory at different experimental trials to avoid participants' adaptation to control only one of the bars for the task of allocating resources.

At the beginning of each trial the two sliders were located at a "null" stimulus mode, namely, at the left most point of the two slider bars. This configuration includes the presentation of a zero cost image and a zero cost audio to the subjects. A grey image was utilised as a zero cost image while a silent track was the zero cost audio track. Grey images are frequently used in experimental studies as a mean of neutralising participant's eyes before the next experimental trial [ITU12]. The two thumbs always reset to the beginning of the bars when the user goes to the next trial. This is not the case when the trial includes the distribution of budget B_5. In that case, both thumbs start from an audio-visual level that has theoretical cost equal to $B_5 - B_4$ and corresponds to a medium quality audio and image. Using this thumb configuration the user was able to start exploring quality levels from the beginning before deciding the desired quality.

The movement of the two thumbs is independent at the beginning until the sum of the costs at the selected audio-visual levels exceeds the budget given for the trial for the first time. In that case, the last moved thumb is placed at the correct location so as the costs sum up to the total budget and the two thumbs become dependent on one another. The dependency of the sliders remains until the user decides to move on to the next trial. Participants become aware of the thumb dependency when they notice the synchronous movement of both slider thumbs.

The transition from independent to dependent GUI controls (thumbs) was decided on purpose so as the user can realise the constraint of the finite budget amount in VE computations. If the thumbs were always independent of one another, the task of resource allocation would be pointless as the users would have the option

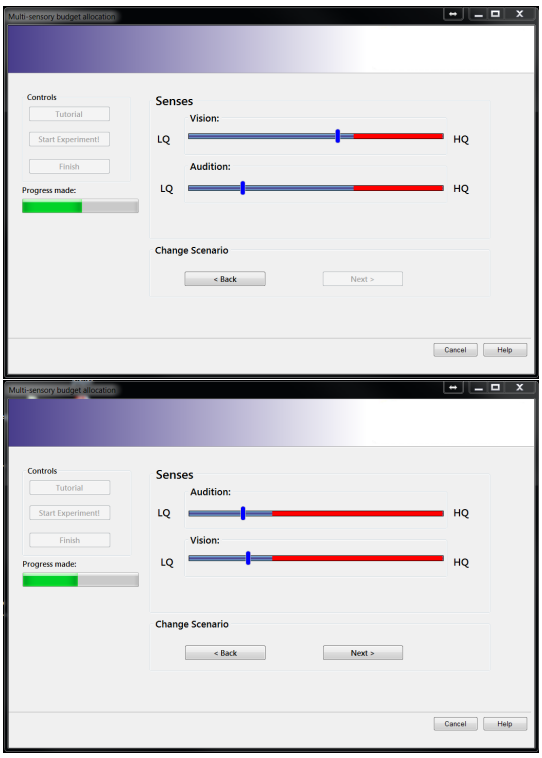


Figure 7.3: Snapshots of the experimental software used including the GUI controls (buttons, slider thumbs, progress bar, etc.). **Top:** An instance of the budget size B_3. **Bottom:** An instance of the smallest budget size B_1.

to maximise both audio and visual qualities at the same time. In that case the sum of the audio-visual costs would never be constant and the budget size could take values other that the ones given in table 7.1. This means, that the *budget* would not be an independent variable. Furthermore, the "next" control button remains disabled as long as the two sliders are still independent and is automatically enabled when they become dependent for the first time. This way, the user had to allocate all the available budget before going to the next trial.

7.3.3 Participants

A total of 35 participants, 18 men and 17 women with ages between 19 and 53 years and from various academic disciplines took part in the experiment. The average age of all the participants was 31.3 years old. All the participants had normal or corrected to normal vision and reported no hearing problem. The sample of participants was opportunistic as it was comprised of university students and academic staff. All the subjects participated voluntarily in the study and received no payment

for their participation.

7.3.4 Procedure

Each participant undertook every possible combination of budget size and scenario for a total of 25 experimental trials. All the experimental sessions were conducted during normal working hours (08:00 am to 17:00 pm) and the length of the experimentation procedure was approximately 20 minutes. The tutorial scenario (Bathroom) was combined with all budgets for five test trials where the users could familiarise themselves with the GUI, the budget configurations and the task of allocating the resources before the main experimental session. There was no constraint on the training time and participants could repeat any of the test trials as many times as they thought necessary. Before the main session, the participants were asked if they thoroughly understood the task of allocating the available resources to the two senses in order to create a satisfactory virtual experience based upon what they saw and heard.

For the experimental study ***E1*** all the necessary ethical considerations have been made. Specifically, all the participants received a Participant Information Leaflet (PIL) before agreeing to take part to the experiment. The PIL explains the procedure and the objectives of the experiment. All the participants had the right to withdraw during or after the end of the experimentation procedure. In case of a withdrawal, all the collected experimental data were destroyed. Furthermore, the ethics documentation includes what happens in case any of the participants had possible side effects, risks, discomforts or complains before, during or after the end of the experiment. Appendix D gives the full contents of the PIL and presents the invitation email for participating to the study. Finally, appendix D includes the ethical approval letter granded for commencing this experimental study and the others that follow in the next chapters.

7.4 Results

In this section the results of the experimental study are presented. The percentage devoted to graphics over the total budget is considered for statistical analysis. Audio allocation is deterministically given from the graphics allocation percentage. Figure 7.4 gives a graphical summary of the experimental data.

As budget size increases, vision allocation percentage decreases proportionally. In other words, vision quality is considered critical when limited resources are given while this trend gradually changes when more resources become available to

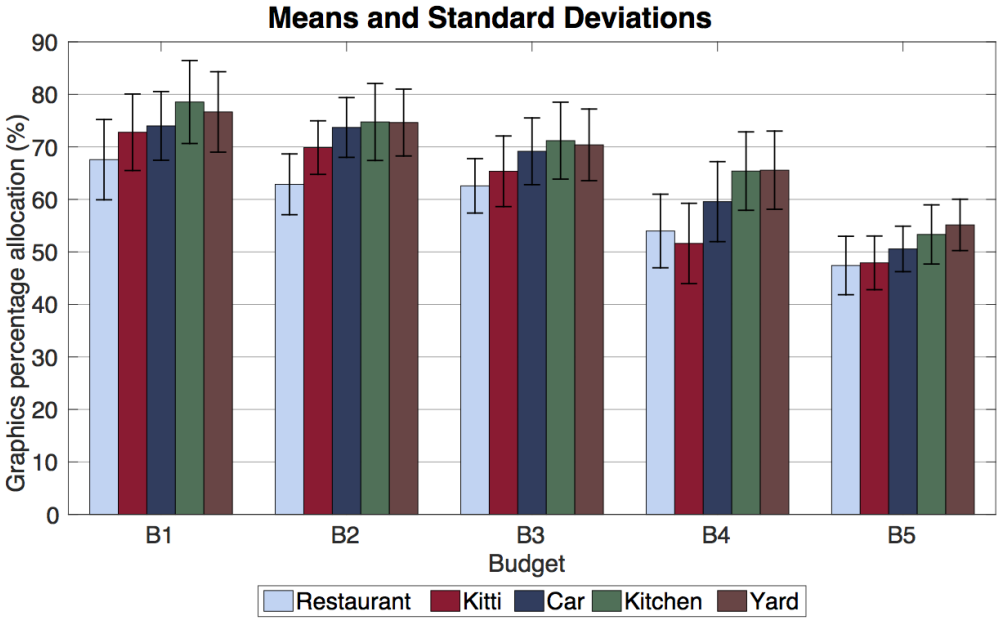


Figure 7.4: Means and Standard deviations for the graphics allocation percentage across all five budgets and experimental scenarios. The scenarios are denoted with differently coloured bars.

the users. At higher budget sizes participants appear to prefer a more balanced distribution of resources for the two senses. Figure 7.4 also shows how graphics allocation percentage varies at different experimental scenarios for a given budget. Vision allocation percentages for Kitti and Restaurant were generally lower than the other three scenarios, indicating a preference for allocating less visual resources in favour of higher aural quality improvements for these scenarios.

The analysis of the data included the application of a 5 (*budget*) $\times$ 5 (*scenario*) factorial design analysed with repeated measures ANOVA as every participant was tested at all possible experimental conditions. This design allows to separate the variability due to the individual subjects from the variation that is explained due to the effect of the two independent variables.

For the overall effect of budget, Mauchly's test showed a violation of sphericity against the budget size ($W(9) = 0.149, p < 0.05$) so Greenhouse-Geisser correction with $\epsilon = 0.56$ was applied. The results revealed that the independent variable *budget* was statistically significant, $F(2.27, 77.2) = 37.5$ and $p < 0.05$. The overall

effect of the *scenario* was also found significant, $F(1.87, 63.6) = 7.27$ and $p < 0.05$; spericity test was violated ($W(9) = 0.154, p < 0.05$) and Greenhouse-Geisser correction with $\epsilon = 0.46$ was applied. The mutual interaction of *budget* and *scenario* did not yield statistical significance, $F(16, 544) = 37.53$ and $p > 0.05$ suggesting that their common interaction does not affect a user's allocation preferences. This can be explained due to the process of normalisation as the scenario complexity does not depend anymore on the budget devoted for its computation. The interaction *budget* $\times$ *scenario* did not violate the assumption of sphericity (via Mauchly's test with $p > 0.05$).

Contrast comparisons between groups of budgets were conducted using post-hoc t-tests to reveal which pairs were statistically significant. All the p-vales were adjusted using Bonferroni correction at a significant level $\alpha = 0.05$. Table 7.2 summarises groups of budgets that are not statistically significant for each scenario and across all the scenarios. Budgets with no significant differences are grouped together. These results indicate that at very small budget sizes, participants follow a similar allocation strategy while as the budget size increases, the budget sizes are significant different indicating their effect on distributing resources.

Scenario	**Budget size**				**p−value**
Restaurant	B_1	B_2 B_3	B_4	B_5	< 0.05
Kitti	B_1	B_2 B_3	B_4	B_5	< 0.05
Car	B_1	B_2 B_3	B_4	B_5	< 0.05
Kitchen	B_1	B_2 B_3	B_4	B_5	< 0.05
Yard	B_1	B_2 B_3	B_4	B_5	< 0.05
All	B_1	B_2 B_3	B_4	B_5	< 0.05

Table 7.2: Contrast comparisons between budgets at every scenario and across all the scenarios. Budgets with no significant differences are grouped together. The level of significance for the hypothesis tests was $\alpha = 0.05$.

Contrast comparisons between groups of scenarios for every budget and across all the budgets were also conducted. These are shown in Table 7.3. These contrasts were used to see if there are significant differences between groups of scenarios. The results indicate that the scenarios Restaurant and Kitti were found significantly different from each one of the other three scenarios presented for the majority of the budgets, meaning that participants generally tend to prefer a significant different resource allocation methodology in these two scenarios (better audio quality).

The human ear is most sensitive to the range 1-4 kHz, which is the range of typical human speech [Gav12]. Motivated by the auditory system's tendency to attend to speech above other sounds [Dar08], an analysis was conducted to determine the size of the difference between those scenarios that had some element of human speech {Restaurant, Kitti} with all the other scenarios. A check was conducted to see if there is significant difference amongst the groups {Restaurant, Kitti} and {Kitchen, Car, Yard} across all the budgets. This comparison can be implemented using the following contrast:

$$L = \sum_{i \in \{\text{scenario}\}} \alpha_i \mu_i, \tag{7.1}$$

where $\alpha_{\text{Restaurant}} = \alpha_{\text{Kitti}} = \frac{1}{2}$ and $\alpha_{\text{Kitchen}} = \alpha_{\text{Car}} = \alpha_{\text{Yard}} = -\frac{1}{3}$ and μ_i denotes the mean allocation for the scenario i. Conducting the hypothesis testing :

$$H_0 : L = 0$$
$$H_1 : L \neq 0,$$

it was found that at significance level $\alpha = 0.05$, H_0 is rejected suggesting that the two groups are significant different ($F(4, 874) = 5.092, p = 0.015 < 0.05$).

Budget	**Scenarios**	**p−value**
B_1	Restaurant | Kitti Car | Kitchen Yard	< 0.05
B_2	Restaurant | Kitti | Car Kitchen Yard	< 0.05
B_3	Restaurant Kitti | Car Kitchen Yard	< 0.05
B_4	Restaurant Kitti | Car | Kitchen Yard	< 0.05
B_5	Restaurant | Kitti | Car | Kitchen | Yard	< 0.05
All	Restaurant Kitti | Car Kitchen Yard	< 0.05

Table 7.3: Contrast comparisons between scenarios at every budget and across all the budgets. Scenarios with no significant differences are grouped together. The level of significance for the hypothesis tests was $\alpha = 0.05$.

Moreover, the mean allocation percentage was computed for every one of the five budgets. This corresponds to a resolution and sampling rate that is shown in Figure 7.5. It is clear that as the budget increases, humans prefer a balanced distribution of resources. This is reflected by the fact that visual quality improvements are implemented in very small steps at large budget sizes because the quality is already high. On the other hand, people start favouring better aural stimuli when their budget permits such decision.

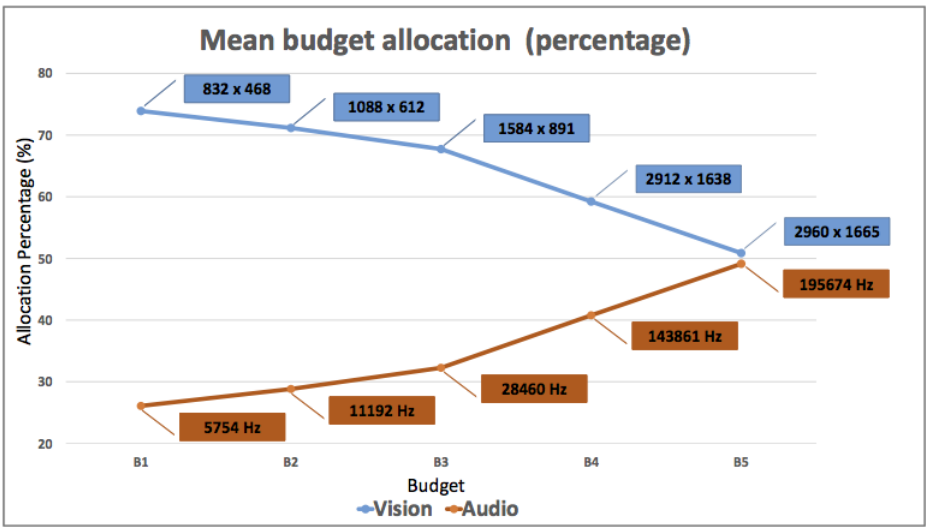


Figure 7.5: Resolution and sampling rate values that correspond to the average allocation percentages for each budget and across all experimental scenarios.

7.5 Estimation models

In order to be able to make use of the obtained results in actual applications, a statistical model was designed that takes the computational budget and scenario type as inputs and provides an audio-visual ratio estimation. The estimations of this model are compared with a more generic statistical model that does not take any information about the scenario type. A second experiment was conducted with a new set of participants to validate the proposed models against actual participants' allocation preferences.

7.5.1 Models

The data obtained from the first experiment was used to construct two regression models that could estimate allocation percentages for the two senses. The first model ($\mathbf{M_1^{E1}}$) takes into account just the budget size while the second ($\mathbf{M_2^{E1}}$) depends both on the budget and the type of the scenario. The construction of $\mathbf{M_2^{E1}}$ was motivated by the results discussed in the previous section, whereby allocation strategy is significantly affected when a scenario includes human voice {Restaurant, Kitti} than when it does not, therefore, the allocation predictions given by model

$\mathbf{M_2^{E1}}$ depend on the input scenario type. The two models are given in the following:

$$\begin{aligned} \mathbf{M_1^{E1}}: &\quad \hat{Y_1} = \hat{\beta_0} + \hat{\beta_1} \cdot \text{budget} \\ \mathbf{M_2^{E1}}: &\quad \hat{Y_2} = \hat{\gamma_0} + \hat{\gamma_1} \cdot \text{budget} + \hat{\gamma_2} \cdot \mathbb{1}_{\text{H}}, \end{aligned} \tag{7.2}$$

where $\hat{Y_1}$ and $\hat{Y_2}$ are the graphics allocation estimations and $\hat{\beta_0}, \hat{\beta_1}$ and $\hat{\gamma_0}, \hat{\gamma_1}, \hat{\gamma_2}$ are the least squares regression estimates of $\mathbf{M_1^{E1}}$ and $\mathbf{M_2^{E1}}$ respectively. The indicator function $\mathbb{1}_{\text{H}}$ takes the unity value when the presented audio stimuli includes human voice (e.g. Restaurant, Kitti) and zero otherwise.

A power (Box-Cox) transform for the power parameter λ was conducted for every one of the two models to examine if transformation of the response variable (allocation) is needed. In both cases the resulting confidence interval contained the value $\lambda = 1$ indicating that no transformation of the response variable was necessary.

Pairs of hypothesis tests for the unknown parameters of the regression models were performed to find their statistical significance. These hypothesis tests are of the following form:

$$\begin{aligned} &H_0 : \beta_i = 0 \text{ vs } H_1 : \beta_i \neq 0, \quad i \in \{0, 1\} \\ &H_0' : \gamma_j = 0 \text{ vs } H_1' : \gamma_j \neq 0, \quad j \in \{0, 1, 2\}. \end{aligned} \tag{7.3}$$

The results show that all the regression parameters are statistically significant (reject H_0 and H_0') for every i and j suggesting that all the model coefficients are important for predicting audio-visual allocation. Specifically, the parameter γ_2 for the scenario type was found to be statistically significant with $t(\text{df} = 873) = 5.53$ and $p = 4.2 \cdot 10^{-6} < 0.05$.

Table 7.4 summarises the regression coefficient estimations and the coefficients of determination for both statistical models. Coefficients of determination (R^2 and R^2_{adj}) give an indication of how much variance in the resource allocation (dependent variable) can be predicted from the budget and scenario (independent variables). Model $\mathbf{M_2^{E1}}$ gives higher R^2 and R^2_{adj} values, indicating a better fitting of the model to the data obtained from the experimental procedure described in the previous section.

7.6 Validation

A second experimental study was conducted to validate the proposed statistical models. The two models have subtle differences and it was desirable to see how the

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	R^2	R^2_{adj}
$\mathbf{M_1^{E1}}$	73.68	-17.83	$*$	$*$	$*$	0.58	0.57
$\mathbf{M_2^{E1}}$	$*$	$*$	76.60	-17.83	-7.31	0.63	0.61

Table 7.4: Estimates of the regression coefficients for models $\mathbf{M_1^{E1}}$ and $\mathbf{M_2^{E1}}$. The respective coefficients of determination are also given for both statistical models (the higher the better).

presence of the extra term $\hat{\gamma_2} \cdot \mathbb{1}_{\text{H}}$ affects the prediction compared to real allocation preferences. This experimental study included the use of untested budget sizes and an untested scenario to investigate performance of the two models in new conditions.

7.6.1 Design

The general design of the experiment follows on from the experimental procedure introduced in sections 7.2 and 7.3. Three scenarios were used for this experiment: Restaurant, Yard and Bathroom, while the Kitti scenario was selected for the training session. It is worth noting that the Bathroom scenario was only used as an introductory tutorial scenario in the previous experiment thus it was utilised for the first time for the resource allocation task. Also both statistical models were built without experimental data coming from this scenario, thus it is of importance to see how the models perform based on this selection as input.

Three test budgets were used in the experiment, each one different from the previously used budgets. The budgets were chosen to lie in the mid-point between B_1 and B_2, B_2 and B_3 and B_3 and B_4. All materials, both at software and hardware level, were prepared in the same way as the first experiment. The procedure was also the same except for the different number of experimental trials. All the participants were tested against all three scenarios and budgets resulting in 9 experimental trials in the main experimental session and 3 trials in the training session. None of the participants in the second experiment had taken part in the first experiment. It was important to use people who were not familiar with the task so as to get unbiased experimental data for the comparison. A total of 10 people, 7 men and 3 women, participated and the average age of all the participants was 28.2 years old. All subjects participated voluntarily and had normal or corrected to normal vision and reported no hearing deficiency.

7.6.2 Results

Figure 7.6 depicts the regression lines for the two models along with the means and standard deviations of the data obtained from the validation experiment. A log-spacing along the x-axis was used to better illustrate the model predictions and the experimental data. In all three cases, $\mathbf{M_2^{E1}}$ performs better with respect to the experimental data obtained from the validation experiment.

Across all three scenarios, the average error of absolute difference for $\mathbf{M_1^{E1}}$ is 3.73% and for $\mathbf{M_2^{E1}}$ is 0.69%. For the previously untested Bathroom scene, results are largely comparable with 3.12% for $\mathbf{M_1^{E1}}$ and 0.59% for $\mathbf{M_2^{E1}}$. The error for model $\mathbf{M_1^{E1}}$ is relatively small on the whole, but compared to $\mathbf{M_2^{E1}}$ it underestimates the graphics allocation for the Yard while it overestimates in the case of the Restaurant scenario. Crucially, even though the models were not constructed using data from the Bathroom scenario the results are accurate, broadly validating the models.

7.7 Discussion

The results of the first experiment yielded a number of potentially interesting findings related to how participants chose to allocate resources. The first relevant finding is that, overall, as expected, under our experimental conditions there was a preference for improving the quality of graphics over the quality of audio. This confirms the original experimental hypothesis as stated in section 4.3.1 and is also supported by previous research by Posner *et al.* [PNK76] and Colavita's study [Col74]. Furthermore, a second relevant finding is that the allocation is dependent on the amount of resources available; when the amount of available resources is increased there is an increase in the amount of resources dedicated to audio. This increase is clear and is statistically significant. This indicates that participants become more concerned with the quality of the acoustic stimulation when paired with higher quality visual stimuli.

The scenarios also demonstrate a significant difference between them. It is presumed that allocation may be scenario dependent and there are indications that this may be the case. As expected [Gav12], the grouping of the scenarios which include human voices provided significant differences compared to those that do not, with participants allocating more resources to audio in the former. The scenario familiarity was not taken into account for this experimental study. Future work will investigate participants' allocation preferences when the scenario is known.

The validation experiment provides evidence that the presented models can be used in general rendering frameworks. Model $\mathbf{M_1^{E1}}$ provides relatively good

results and $\mathbf{M_2^{E1}}$ even stronger as long as the scenarios can be correctly categorized. Rendering frameworks can consult the models for a given computational load or achievable frame rate and identify how much of that time is to be dedicated to graphics and acoustics.

7.7.1 Limitations

Further work with more scenarios is required to provide an understanding of whether the scenarios can be classified in the manner described above, or what other characteristics can be used to further predict correct utilisation of resources. A possible limitation of the work presented in this chapter may be due to the static nature of the graphics. It is important to verify whether these observations hold when videos or interactive environments are used for the visuals.

The normalisation of the functions when assigning costs to the visual/auditory stimuli might also be considered as a limitation in terms of the generalisation of the results. However, the alternative of working with real, measured, computational budgets requires fixing the acoustic and graphical rendering to particular, distinct algorithms. As is known, image and audio rendering algorithms can vary significantly in terms of computation, possibly by several orders of magnitude; for instance, path traced and rasterised images. The selection of such methodology would be difficult to generalise further and it would limit the scope of the current work.

The goal of this work was to identify the right amount of resources to dedicate to each sense from an available computational budget. In this respect, the proposed model and the experimental results can be used as an initial guideline for audio-visual resource allocation.

7.8 Summary

This chapter presented a subjective experiment where participants ($N = 35$) allocated five fixed resource budgets across visual and acoustic stimuli. In the experiment, increasing the quality of one of the stimuli directly meant that the quality of the other should be decreased such that the cost of the two displayed stimulations remains constant and equal to the user's allocated budget. The results demonstrate that participants allocate more resources to graphics, however as the computational budget is increased people clearly tend to prefer a balanced allocation strategy between vision and audio. The results found confirm the original hypothesis as stated in section 4.3.1. Graphics, in general, dominate the resource allocation task as people prefer more resources to be distributed for better visual quality. Based on the

results, an audiovisual quality prediction model is proposed and successfully validated using new budgets and a combination of new and old experimental scenarios.

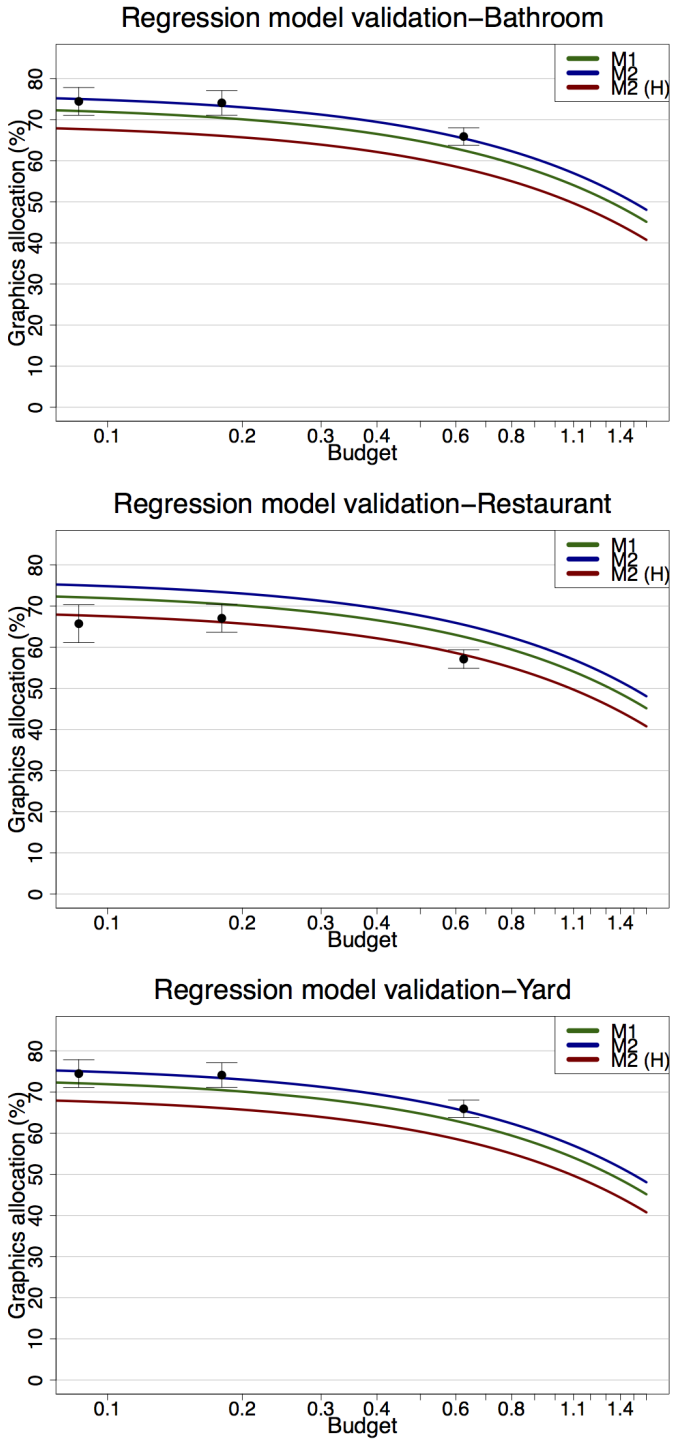


Figure 7.6: Means and standard deviations are depicted in black colour. The green curve shows the predictions of the model $\mathbf{M_1^{E1}}$ while the blue and red coloured curves show the estimations of the model $\mathbf{M_2^{E1}}$ when a human voice is absent or present in the scenario respectively. Log-spacing was applied to the x-axis for better visualisation.

Chapter 8

Olfactory pipeline and materials

8.1 Introduction

This chapter describes the methodology followed for the simulation of smell propagation phenomena in VEs and the delivery of smell impulses to the users using an olfactory display. Smell transport simulation in VEs is implemented via Computational Fluid Dynamics (CFD). CFD can be used to model the physical processes of diffusion and convection that govern the transport of an odorous gas in an environment. During the simulation stage, smell concentration values are recorded at different areas of the domain using virtual sensors. These physically accurate concentration values are subsequently directed to an olfactory display that is used to deliver the smell impulses to the participants at specific temporal accuracy.

This chapter also investigates how the concentration of a propagating smell varies at specific locations of a VE when a better spatial discretisation level of the computational domain is applied prior to the simulation. Better spatial discretisation is a commonly used technique to increase numerical accuracy in CFD problems and is achieved using higher tessellation fineness of the computational domain into finite volumes. These finite volumes compose the computational domain where the smell is virtually propagated.

This chapter is organised as follows. Firstly, the mathematical background for simulating odour transport in VEs using CFD is presented in section 8.2. Secondly, section 8.3 discusses the set-up used for simulating odour transport in four VE case studies and investigates the effect of different discretisation levels at the concentration results captured using virtual probes located at carefully chosen locations in the VEs. Finally, section 8.4 discusses limitations of the work described throughout the chapter and gives possible directions for future work. The content

of this chapter is the basis for the implementation of the experimental studies ***E2*** and ***E3*** of this thesis.

8.2 Smell transport simulation background

The problem of smell transport can be approached either from the microscopic or the macroscopic points of view. Its complexity varies depending on the problem requirements and the level of accuracy needed. All the smell transport applications throughout this thesis approach the problem macroscopically using CFD techniques and specifically the *Finite volume method* on descritised geometrical domains. The selection of CFD allows the simulation of both diffusion and convection phenomena at the same time, thus giving accurate estimations of odour concentration at any point of the VE.

8.2.1 Simulating smell transport at micro level

At a microscopic level, the odorous particles that compose the odour perform random movement into a medium (e.g. air). This behaviour characterises the physical process of diffusion and was described earlier in section 2.4.1. The particle random motion can be mathematically formulated using a *Brownian motion* [Bro27] which describes how the movement of a smell particle is affected by the continuous collisions with the particles that constitute the medium.

The 3D Brownian motion is a stochastic Markov process ($\mathbf{B}_t = (X_t, Y_t, Z_t),\ t \geq 0$) where every coordinate is a real-valued Brownian motion that evolves in time and space independently of the other two. The initial position of the Brownian motion is given at time $t = 0$ and is B_0. The Brownian motion has transition kernel given by:

$$P_s^{\mu}(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d} P_s^{\mu_i}(x_i, y_i) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2s}(x_i - \mu_i - y_i)^2}, \tag{8.1}$$

where $P_s^{\mu}(\mathbf{x}, \mathbf{y})$ is the Probability Density Function (PDF) that characterises the particle movement from position $\mathbf{x}$ to position $\mathbf{y}$ in space within a time interval of length s. The letter d is used to denote the dimension of the Brownian motion while the letter μ is known as its *drift* and determines the line around which the process performs its random fluctuations. Figure 8.1 shows the effect of different drift values in a 1D Brownian motion.

If we know that the process at time t is at a point $\mathbf{v} = (x, y, z)$ in space, i.e. $\mathbf{B}_t = \mathbf{v}$, then we can find the probability at time $t + s$ the process to be inside a

region $A \subset \mathbb{R}^3$. This probability can be given as:

$$\mathbb{P}(\mathbf{B}_{t+s} \in A \mid \mathbf{B}_t = \mathbf{v}) = \int_A P_s^{\mu}(\mathbf{x}, \mathbf{y}) dA. \tag{8.2}$$

A Brownian motion can be implemented using a random walk that moves in space according to the positions that are sampled from the transition kernel given in equation 8.1. For example, for the Brownian motion $\mathbf{B}_t$, when we know that its component X_t is at position x_1 at time t_1 (i.e. $X_{t_1} = x_1$) at the next discrete time step its position can be written as:

$$X_{t_2} = x_1 + \alpha$$

where α is sampled by the PDF given in formula 8.1 with parameters μ_x and $s = t_2 - t_1$. Figure 8.1 shows an odour diffusion simulation using Brownian motion for a total of 500 molecules.

Brownian motion is easy to implement while it can simulate smell diffusion in real time. One of its main drawbacks is that it cannot be used to model any convective flows. As a result, someone needs to assume that the domain of interest has no external disturbances; i.e. it is a sealed room. In this thesis, odour transport has been implemented using CFD for permitting both diffusion and convection modelling of the fluid mass into the geometrical domain. For this purpose, a discretisation of the governing equations and the geometrical domain was required. In the following, the CFD pipeline is presented while section 8.3 presents the simulation set-up for the domains of interest.

8.2.2 Simulating smell transport at macro level

Any fluid can be considered a medium where at higher macroscopic scales the structure and molecular properties of its individual particles are negligible compared to the overall behaviour of the fluid. From the macroscopic point of view the fluid transport can be described using quantities such as velocity ($\mathbf{u}(x, y, z, t)$), density ($\rho(x, y, z, t)$), pressure ($p(x, y, z, t)$) and temperature ($T(x, y, z, t)$) as well as their derivatives at any given position and time. The macroscopic approach considers the average values of these quantities over a massive number of particles and studies their variations within bounded volumes that compose the domain of interest.

The fluid dynamics pipeline for smell transport begins with the discretisation of the problem's geometrical domain. Domain discretisation includes the tessellation of its interior and boundary surfaces into a finite number of cells. Central in the

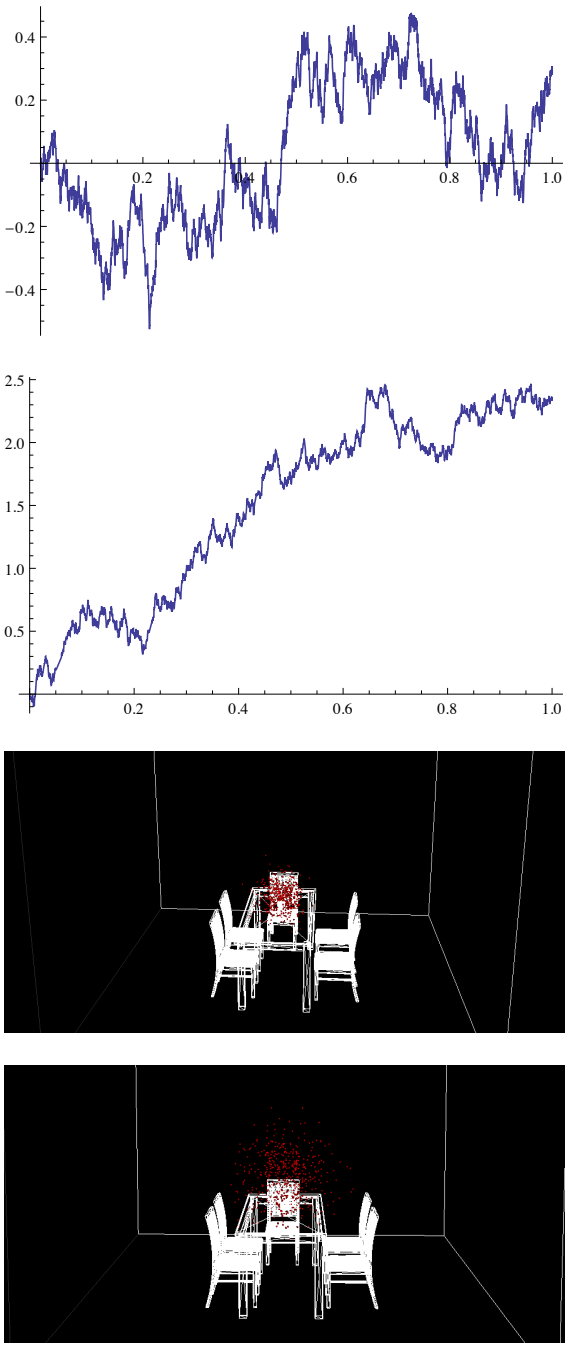


Figure 8.1: **First (Top):** 1D Brownian motion with drift $\mu = 0$, **Second:** 1D Brownian motion with $\mu = 2$. The random process fluctuates around the line $y = 2x$, **Third:** Simulation of 200 diffusing particles using 3D Brownian motion. Snapshot after 30 sec of virtual time passed. **Fourth (Bottom):** Same as left, snapshot after 723 sec of virtual time passed.

fluid dynamics pipeline is the discretisation of the governing equations that describe the physics of the fluid transport. The finite volume method is used to discretise the partial differential equations into a set of linear equations that compose an algebraic system.

Before proceeding to the solution of the system, boundary conditions that describe the physical requirements are set. Boundary cells are used to impose the physical requirements of the problem. For instance, they can be used to set up inlet or outlet conditions where the mixture of air and smell enters or leaves the domain respectively. The system of linear equations is solved to yield the distribution of the variables (e.g. velocity, pressure, odour concentration, etc.) at the centres of the mesh cells and across a range of consecutive time steps.

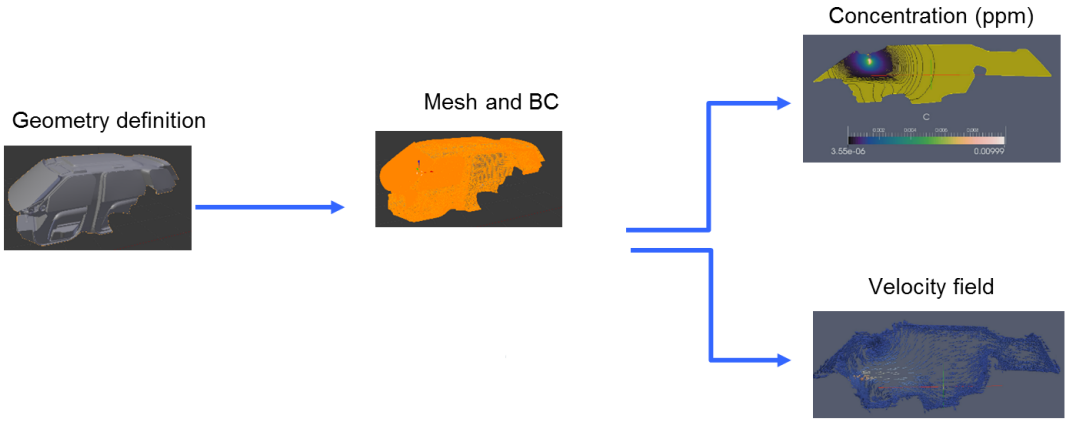


Figure 8.2: Olfactory simulation pipeline using CFD [VM07]: The geometry is used to produce a tessellation of the domain. Boundary conditions and discretisation of the governing equations are also applied before the computation. The result is the concentration, C, and a velocity vector field, $\mathbf{U}$ at the centres of the polyhedra that compose the mesh. All the images were generated using the ParaView visualiser and represent the car CAD model used in this chapter [Aya15]

.

8.2.3 Discretisation of the geometrical domain

The discretisation of the solution domain includes its subdivision into smaller elements known as *Control Volumes* (CVs). CVs are convex polyhedra that have planar faces. Also they should completely fill the domain and cannot overlap. A typical example of a control volume is given in Figure 8.3. The centre of the V_d control volume is denoted as P and its volume as V_p. The vector $\mathbf{d}$ connects the centres of two neighbouring CVs with centres P and N respectively. Volume faces

are denoted as f while the normal vector at the center of the face points outwards and it is notated as $\mathbf{S_f}$. The $\mathbf{S_f}$ vector has magnitude equal to the area of the face f.

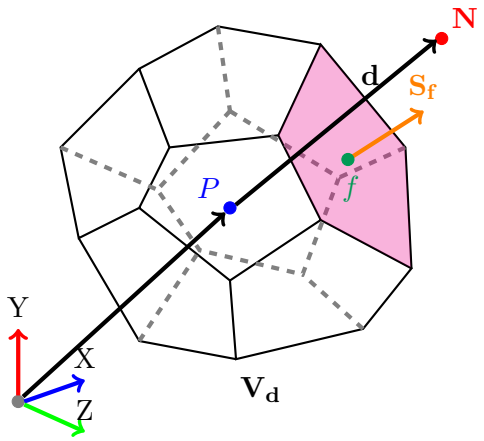


Figure 8.3: An example of a polyhedron control volume. The figure was adopted by the OpenFoam documentation guide [WTJF98].

The center P of the CV as well as the center of any of its faces f can be given using the equations:

$$\begin{aligned} \int_{V_p} (\mathbf{x} - \mathbf{x_P})\, dV = 0 \\ \int_{S_f} (\mathbf{x} - \mathbf{x_P})\, dS = 0 \end{aligned} \tag{8.3}$$

A face can either be internal and shared between two cells or located at the boundary of the domain.

8.2.4 Discretisation of the governing equations

The initial step for every fluid transport simulation is the governing equations that describe the physical processes of the problem. These are known as the *Navier-Stokes* differential equations and can be derived using mass, momentum (Newton's second law) and energy (first law of thermodynamics) conservation principles (see [VM07] for more details). These equations make use of the gradient (∇) and divergence ($\nabla\bullet$ or div$(\cdot)$) differential operators which are explained in Appendix C.

The governing equations can be written as [VM07]:

$$\text{Continuity: } \frac{\partial \rho}{\partial t} + \nabla \bullet (\rho \mathbf{U}) = 0,$$

$$\text{X-momentum: } \frac{\partial (\rho u)}{\partial t} + \nabla \bullet (\rho u \mathbf{U}) = -\frac{\partial p}{\partial t} + \nabla \bullet (\mu \nabla u) + S_{Mx},$$

$$\text{Y-momentum: } \frac{\partial (\rho v)}{\partial t} + \nabla \bullet (\rho v \mathbf{U}) = -\frac{\partial p}{\partial t} + \nabla \bullet (\mu \nabla v) + S_{My}, \tag{8.4}$$

$$\text{Z-momentum: } \frac{\partial (\rho w)}{\partial t} + \nabla \bullet (\rho w \mathbf{U}) = -\frac{\partial p}{\partial t} + \nabla \bullet (\mu \nabla w) + S_{Mz},$$

$$\text{Internal energy: } \frac{\partial (\rho i)}{\partial t} + \nabla \bullet (\rho i \mathbf{U}) = -p \nabla U + \nabla \bullet (k \nabla T) + S_i,$$

where the vector-valued function $\mathbf{U} = (u, v, w)$ is the fluid's momentum and the scalar-valued functions p and ρ are the pressure and density of the fluid respectively. The letter i denotes the internal energy of the fluid and is measured in Joules. The constant μ is the dynamic viscosity and expresses how easily a fluid can be deformed under the effect of shear stresses that act upon it. Dynamic viscosity is a physical property of the fluid and is measured in units of kg/(s·m). It expresses how viscid a fluid is. For standard temperature conditions the viscosity of air is 1.983×10^{-5} kg/(s · m) whilst the viscosity of different gases ranges in that order of magnitude. It should be noted that viscosity increases for higher temperatures in the environment. Reynolds law [Rey86] describes this relation as:

$$\mu(T) = \mu_0 e^{-bT}, \tag{8.5}$$

μ_0, b are constants and T is the temperature. The terms $S_{Mx}, S_{My}, S_{Mz}, S_i$ are known as source terms of the momentum and energy equations and are used to describe viscous stresses and other body forces that are exerted in the fluid during its transport. The constant k expresses how much energy in the form of heat is conducted per unit meter and temperature and is known as *Thermal conductivity* of the fluid. It is measured in W/(m · K).

The five equations given in 8.4 involve seven variables that need to be found, namely, u, v, w, ρ, p, i and T. This system of equations is not closed and two more equations are required. These are known as *Equations of state* and connect some of

the variables of interest as such:

$$p = \rho RT \qquad\qquad i = C_v T, \tag{8.6}$$

The constant R is the universal gas constant and has value 8.314 $J \cdot K^{-1} \cdot \text{mol}^{-1}$. The second constant C_v is the specific heat of the gas when its volume remains unchanged and expresses how much heat is needed to increase temperature by one Kelvin (units: Joule/K) in the fluid. The use of these state equations presupposes that the gas is under thermodynamic equilibrium meaning that possible thermodynamic changes, for instance, temperature or pressure variations happen so fast that the fluid's state immediately adjusts to these changes [VM07].

Navier-Stokes equations can be further customised depending on whether the fluid flow is *Compressible* or *Incompressible*. Incompressible flows assume that the density ρ remains approximately constant during the flow at any point of the domain. This assumption slightly simplifies the system of equations as one less variable needs to be specified. However, in cases where external factors (e.g. changes in temperature or pressure) cause density variations the flow is compressible. For example, in a smell transport problem where the room temperature is constant and the boundary conditions do not impose any temperature variations the flow is incompressible. In case that the room contains sources of heat, these create convective flows (see section 2.4.1) that transfer heat and subsequently affect temperature and density at different regions of the domain. This is a typical example where a compressible flow needs to be modelled.

Modelling turbulence

The majority of the fluid flows that are found in nature present an unstable behaviour that is responsible for the transition from *laminar*, where the fluid flows in straight trajectories, to *turbulent* where the fluid performs vortices (*eddies*) of various sizes in a random fashion. A turbulent flow is characterised by a continuous variation of the velocity and pressure over a wide range of time and length scales at the various eddies that are created during the flow [LMC92]. A flow is characterised as turbulent if it has high Reynolds number. This is defined as:

$$Re = \frac{\rho \mathbf{U} L}{\mu}, \tag{8.7}$$

where μ is the dynamic viscosity of the fluid and ρ its density. The symbols $\mathbf{U}$ and L are used in this context to denote the characteristic velocity of the fluid and the characteristic length of the domain that the fluid flows into. The characteristic

velocity expresses the maximum velocity that the fluid can obtain due to the effects of different obstacles present in the domain. The characteristic length expresses the physical length that is covered by the fluid during its flow. For example, when inflating a balloon with helium, the characteristic velocity is the speed of the air flow at the edge of the pump while the characteristic velocity is the balloon's maximum height.

The Reynolds number expresses the ratio between the inertial forces and the viscous forces exerted on the fluid during its transport. Inertial forces, as the name suggests, are attributed due to the momentum of the fluid whilst viscous forces are forces which are developed due to the friction that is developed between the fluid and the geometrical surfaces of the VE. When viscous and inertial forces are comparable in magnitude the Reynolds number is relatively small and the flow is characterised by laminar trajectories otherwise the flow is turbulent. In smell simulation problems, the viscosity of the transported odour is very low (of the order of 10^{-5}) yielding high Reynolds numbers (of the order of 10^4) and therefore the flow of air-odour mixture presents a turbulent behaviour.

There is a plethora of different ways to model and simulate turbulence in fluid flow applications. This is usually achieved through the use of turbulence models. Turbulence models have been developed by collecting experimental data during the observation of physical flows under different environmental conditions. The best known and most frequently used turbulence models are the $k-\epsilon$ and Reynolds stress models [VM07]. The $k-\epsilon$ model, specifically, models the effect of two more variables by considering two more PDEs in the governing equations introduced earlier. These variables are the *turbulent kinetic energy* denoted as k and measured in m^2/s^2 and the second is the dissipation of the turbulent kinetic energy per unit mass denoted as ϵ and measured in m^2/s^3. Study of turbulence in fluid mechanics applications is an active area of research and many models have been proposed in the past for obtaining turbulent effects in fluid flow simulations. The interested reader is referred to Jones and Launder [JL72] and Versteeg and Malalasekera [VM07].

Transport equation

Navier-Stokes equations can be used to predict the velocity, pressure, density and energy of the mixture over time but do not provide any information regarding the concentration of the odour in the air. The *transport equation* is a PDE that is solved along with the governing equations and describes how a scalar quantity such as odour concentration C changes at different areas of the domain and under the combined effect of diffusion and convection (see section 2.4.1). The general form of

the transport equation is given as follows:

$$\underbrace{\frac{\partial \rho C}{\partial t}}_{\text{temporal derivative}} + \underbrace{\nabla \bullet (\rho \mathbf{U} C)}_{\text{convective term}} - \underbrace{\nabla \bullet (\rho D_{sa} \nabla C)}_{\text{diffusion term}} = \underbrace{S_C(C)}_{\text{source term}} , \tag{8.8}$$

where ρ and D_{sa} are the density and the diffusion coefficient of the transported quantity C. Every term of the transport equation is used to model the effect of a different physical process in the odour transport simulation. In particular, convection and diffusion are modeled using the terms $\nabla \bullet (\rho \mathbf{U} C)$ and $\nabla \bullet (\rho D_{sa} \nabla C)$ respectively while the source term $S_C(C)$ is used to capture earth's gravitational field or other body forces (e.g. centrifugal or coriolis forces). Odour concentration in air is subject not only to spatial variation but also to changes over time, therefore, the presence of the temporal term in the transport equation is necessary to capture this behaviour 8.8.

The diffusion coefficient or *mass diffucivity* expresses the rate an odour diffuses into air. It is measured in $\frac{\mathrm{m}^2}{s}$ while its value is affected by the temperature and pressure conditions of the environment. Diffusion coefficients for VOCs can be estimated using mathematical formulas [MM72] or through experimentation procedures [FSQM03]. A frequently used formula for estimating the diffusion coefficient of a VOC into air is given by the *Chapman-Enskog theory* [Cus97]. According to this, a diffusion coefficient of a VOC into air can be computed as:

$$D_{sa} = 1.858 \cdot 10^{-3} \frac{T^{\frac{3}{2}} \cdot \sqrt{\frac{1}{M_s} + \frac{1}{M_a}}}{p \sigma_{sa}^2 \Omega}, \tag{8.9}$$

where the subscript "sa" is used to denote the diffusion rate of one substance (smell) into the other (air). In the above equation, T is the temperature measured in Kelvin, M_s, M_a are the molar masses of the VOC and air respectively. The symbol p is used to denote the pressure while $\sigma_{sa} = \frac{1}{2}(\sigma_s + \sigma_a)$ is the average collision diameter which can be computed using the molecule diameters (σ_1, σ_2). These can be obtained from tabulated data [MM72]. Finally, Ω is known as the temperature dependent collision integral and is used to express energy transport during molecule collisions [MM72].

The transport equation is a second order PDE because of the existence of the Laplacian operator at the diffusion term. For preserving numerical accuracy, the discretisation of the PDEs needs to be at least second order accurate in space and time. For the scalar quantity $C = C(\mathbf{x}, t)$ a linear profile around the cell centres

is assumed as follows:

$$\begin{aligned} C(\mathbf{x}) &= C_P + (\mathbf{x} - \mathbf{x_P}) \cdot \big(\nabla C\big)_P \\ C(t + \Delta t) &= C^t + \Delta t \Big(\frac{\partial C}{\partial t}\Big)^t, \end{aligned} \tag{8.10}$$

where $C_P = C(\mathbf{x_P})$ and $C^t = C(t)$. The above equations are the Taylor expansion (described in Appendix C) of the function C around x_P and Δt respectively. The first term of the remainder error in the first expansion is of the order of $\mathcal{O}\big((\mathbf{x} - \mathbf{x_p})^2\big)$. When the difference $\mathbf{x} - \mathbf{x_P}$ is sufficiently small (less that unity) at a control volume this quadratic term yields second order numerical accuracy as the other terms in the expansion are negligible.

The assumed profile for the function C can be used to compute the integral over a CV:

$$\begin{aligned} \int_{V_P} C(\mathbf{x})\, dV &= \int_{V_P} \big[C_P + (\mathbf{x} - \mathbf{x_P}) \cdot \big(\nabla C\big)_P\big]\, dV \\ &= C_P \int_{V_P} dV + \Big[\int_{V_P} (\mathbf{x} - \mathbf{x_P})\, dV\Big] \cdot \big(\nabla C\big)_P \\ &= C_P V_P, \end{aligned} \tag{8.11}$$

where the second integral at the right hand side of the above equation yields zero because the point P is the center of the CV (see equation 8.3).

The finite volume method is a technique for discretising PDEs over a CV with volume V_P and across a time interval $[t, t+\Delta t]$. The value Δt denotes the time step that is used for discretising the transport equation in the temporal domain. All the functions that are integrated over a CV need to be discretised using the values of the variables at the faces that compose a CV. The discretisation of the governing equations follows exactly the same methodology thus the focus in the following will be only in the transport equation. The first step at the finite volume method is the integration of the transport equation 8.8 with respect to time and space. This dual integration is written as:

$$\begin{aligned} \int_t^{t+\Delta t} \Big[\frac{\partial}{\partial t} \int_{V_P} \rho C\, dV + \int_{V_P} \nabla \bullet (\rho \mathbf{U} C)\, dV - \int_{V_P} \nabla \bullet (\rho D_{sa} \nabla C)\, dV\Big] dt \\ = \int_t^{t+\Delta t} \Big[\int_{V_p} S_C(C)\, dV\Big] dt \end{aligned} \tag{8.12}$$

Discretisation of the transport equation

The discretisation of the transport equation includes the numerical approximation of all the integrals shown in equation 8.12. The first step is to estimate the terms that include integration over a CV and after that the integrals with respect to time. For estimating the integrals that contain differential operators it is necessary to use Gauss's theorem of divergence. This theorem states that the integration of the divergence operator applied to a vector-valued function $\mathbf{g}$ over a CV can be expressed as an integral over the faces that bound the CV. Specifically:

$$\begin{aligned}\int_{V_P} \nabla \bullet \mathbf{g}\, dV &= \oint_{\partial V_P} \mathbf{g} \cdot \mathbf{n}\, dS \\ &= \sum_f \Big(\int_f d\mathbf{S} \cdot \mathbf{g} \Big),\end{aligned} \tag{8.13}$$

where the ∂V_P is the boundary that surrounds the CV and dS is the integration measure over this boundary. The vector $\mathbf{n}$ represents the outward pointing vector normal to the faces that compose the boundary. A frequently used notation is to write $\mathbf{n}dS$ as $d\mathbf{S}$ and express the integrand at the right hand side of the above equation as $\mathbf{g} \cdot d\mathbf{S}$. Also the integral over the boundary can be split into the faces that compose it. This can be seen at the second line of equation 8.13.

The face integral at equation 8.13 can be calculated using the assumed linear profile of the function $\mathbf{g}$ as follows:

$$\begin{aligned}\int_f d\mathbf{S} \cdot \mathbf{g} &= \Big(\int_f d\mathbf{S} \Big) \mathbf{g_f} + \Big[\int_f d\mathbf{S}(\mathbf{x} - \mathbf{x_f}) \Big] \cdot (\nabla \mathbf{g})_f \\ &= \mathbf{S_f} \cdot \mathbf{g_f},\end{aligned} \tag{8.14}$$

where $\mathbf{g_f}$ denotes the value of the function at the centre of a face f of a CV. Again the second integral at the right hand side has zero value because f is the center of the face (see equation 8.3).

Equations 8.14 and 8.13 combined with the result given at equation 8.11 yield a discrete approximation of the divergence of the function $\mathbf{g}$ as follows:

$$\big(\nabla \bullet \mathbf{g}\big)_P V_P = \sum_f \mathbf{S_f} \cdot \mathbf{g_f} \tag{8.15}$$

This last result is the cornerstone for discretising the convection term at the

integrated transport equation 8.12 as follows:

$$\begin{aligned}\int_{V_P} \nabla \bullet \left(\rho \mathbf{U} C\right) dV &= \sum_f \mathbf{S_f} \cdot \left(\rho \mathbf{U} C\right)_f \\ &= \sum_f \mathbf{S_f} \cdot \left(\rho \mathbf{U}\right)_f C_f.\end{aligned} \tag{8.16}$$

The inner product $F = \mathbf{S_f} \cdot \left(\rho \mathbf{U}\right)_f$ is used to indicate the flux of mass through a face f. The same discretisation practice is also applied for the integral of the diffusion term shown in equation 8.12:

$$\begin{aligned}\int_{V_P} \nabla \bullet \left(\rho D_C \nabla C\right) dV &= \sum_f \mathbf{S_f} \cdot \left(\rho D_{sa} \nabla C\right)_f \\ &= \sum_f \left(\rho D_C\right)_f \mathbf{S_f} \cdot \left(\nabla C\right)_f\end{aligned} \tag{8.17}$$

As can be seen from the discretisation of the convection and diffusion terms, It is required to compute values of the transported quantity C and its gradient at the faces of the CV. For this reason, it is necessary to interpolate the values given at the centres of the control volumes and get estimations for the variables of interest at the faces of the CV. The numerical methods used for value interpolation across nodes of a mesh are known as *differencing schemes* and yield different levels of numerical accuracy at the cost of extra computational and storage requirements. There is a vast literature regarding the advantages and disadvantages of the differencing schemes and their importance in CFD simulations. The interested reader can find the theory and practical applications of these schemes in: Jasak's phd thesis [Jas96], Versteeg and Malalasekera [VM07] and Ferziger and Peric [FP02]. In this thesis, the focus will be in explaining the three basic properties of the differencing schemes in the context of odour transport simulation.

One of the basic properties of a differencing scheme is its ability to preserve bounded values for the computed variables. The *Boundedness* of a solution refers to the existence of values M_1, M_2 such that $M_1 \leq C(\mathbf{x}, t) \leq M_2$ for every $\mathbf{x}$ in the computational mesh and time point t. There are differencing schemes that yield unbounded solutions for the variables of interest. It is quite possible that a simulation will give unrealistic values when using schemes that yield unbounded solutions. For instance, negative concentration values C can be produced when using inappropriate differencing schemes.

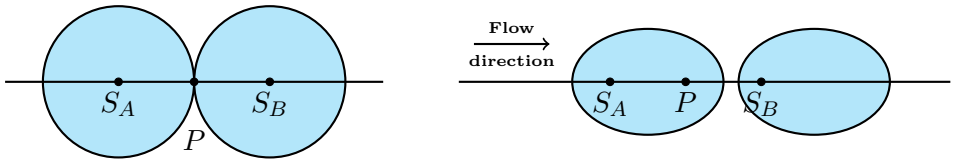
Another basic property is the *Conservativeness* of the transported scalar. According to this property, the amount of a transported quantity that leaves a CV

across one of its faces should be equal to the amount that enters the neighbouring CV though the same face. In other words, the transported quantity needs to be conserved at the CVs that compose the computational mesh. If for any reason there are losses due to incorrect selection of differencing scheme the simulation yields inaccurate results or diverges.

The last property of a differencing scheme is known as *Transportiveness* and refers to the ability to capture both diffusion and convection phenomena without loss of numerical accuracy. A way to see whether the physical process of diffusion or convection dominates at a fluid flow is by calculating the *Peclet* number of the flow as follows:

$$Pe = \frac{FL}{D_{sa}}, \tag{8.18}$$

where F is the mass flux, D_{sa} is the mass diffusivity and L is the characteristic length of the problem. An example of the transportiveness property is given in image 8.4. This Figure depicts a 1D diffusion/convection problem where two sources, namely, S_A and S_B are placed at diametrically opposed positions around a point P. It can be assumed that these are sources of a scalar quantity C (odour sources) for simplicity. In figure (a) the fluid flow (e.g. air flow) is relatively negligible and the two sources diffuse creating regions of higher values of odour concentration C around their positions. In that case the value of C at point P will be equally dependent on both sources and the Peclet number will have a very small value because the flow velocity is almost 0.



(a) (b)

Figure 8.4: **Left:** The fluid transport is governed by diffusion giving a Peclet number close to 0. **Right:** The fluid transport is governed by convection giving a higher value for the Peclet number. The figure was generated based on the theory described in Versteeg's and Malalasekera's book [VM07].

In figure (b) air flows in the right direction and convects the odour to the right. In that case the value of the variable C at the point P is influenced solely by the source S_A while the source S_B has no contribution. The Peclet number gets a

higher value as convection dominates the transport of the quantity C. A differencing scheme that possesses the transportiveness property needs to deal equally well to these cases taking into account the flow direction and the Peclet number.

As far as the source term is concerned, it is modelled depending on the existence of body forces and the various requirements of the problem. Extended guidelines for appropriately modelling the source terms can be found in [SB78]. In the simplest case, the source term can be modelled as a linear function of the variable C as follows [Jas96]:

$$S_C(C) = S_U + S_P C, \tag{8.19}$$

where S_U and S_P are also functions of C. The integral of the source term can be discretised as:

$$\int_{V_P} S_C(C)\, dV = S_U V_P + S_P V_P C_P.$$

Including the discritised formulations of all the integrals over a CV, the transport equation 8.12 takes the following form:

$$\begin{aligned} \int_t^{t+\Delta t} \left[\left(\frac{\partial \rho C}{\partial t}\right)_P V_P + \sum_f F C_f - \sum_f \left(\rho D_{sa}\right)_f \mathbf{S_f} \cdot \left(\nabla C\right)_f \right] dt \\ = \int_t^{t+\Delta t} \left(S_U V_P + S_P V_P C_P\right) dt. \end{aligned} \tag{8.20}$$

The final step in the discretisation process is the approximation of the temporal derivative and the discretisation of both integrals at the left/right hand side of the equation 8.20. In general, the integral of a scalar-valued function C over a time interval $[t, t + \Delta t]$ can be estimated using:

$$\int_t^{t+\Delta t} C(t) = \frac{\left(C^0 + C^n\right)\Delta t}{2}, \tag{8.21}$$

while the time derivative at the cell centre$P$ can be approximated as:

$$\left(\frac{\partial \rho C}{\partial t}\right)_P = \frac{\rho_P^n C_P^n - \rho_P^0 C_P^0}{\Delta t}, \tag{8.22}$$

where $C^0 = C(t)$ and $C^n = C(t + \Delta t)$ are values of the scalar quantity C at different time steps at the face centres of a CV. The equations 8.21 and 8.22 can be derived using the linear profile of the scalar quantity C in time as it is written in equation 8.10. Applying these formulae at the integrals present in equation 8.20 and estimating the temporal derivative using the formula—8.22 one can derive the

final form of the discretised transport equation.

It should be noted here that there are many different temporal differencing schemes for estimating values of the simulated variable at the faces of a CV based on one or more values of the variable at older time steps [Jas96, VM07, FP02].

8.2.5 Systems of linear equations

As pointed out previously, the values of the variable C at the faces of the CVs can be estimated through the use of differencing schemes. These schemes interpolate the values at the centres of the CVs to get numerical estimations at the faces. A final discretised linear equation is written for every node of the mesh and involves the values of the variable C at this node and its neighbours. The collection of all the linear equations gives an algebraic system as follows:

$$\alpha_P C_P^n + \sum_N \alpha_N C_N^n = B_P, \tag{8.23}$$

where C_P^n and C_N^n are the values at the node P and its N neighbours respectively whilst α_P and α_N are the coefficients of these values. The above system of equations can be compactly written in the form:

$$\underline{A} \cdot C = \underline{B}, \tag{8.24}$$

where $\underline{A}$ is the coefficient matrix, C is the vector of the unknown values at the nodes of the CVs and $\underline{B}$ is a vector with constant values. The non-zero values at every row of the coefficient matrix show the neighbouring nodes that are involved for the computation of the variable at the node P. As every CV is not connected with all the other CVs but only with its neighbours, many of the coefficients present at the linear equation are zero meaning that many of the entries of the coefficient matrix are zero. This produces an algebraic system where the matrix $\underline{A}$ is sparse.

The system of linear equations can be solved either using *Direct* or *Iterative* methods. Direct methods include the application of Gaussian elimination that gives the exact solution at algorithmic complexity $\mathcal{O}(N^3)$ where N is the size of the coefficient matrix $\underline{A}$. The complexity of the direct methods is prohibitive especially for large computational meshes where the coefficient matrix $\underline{A}$ is very large [FP02].

The usual methodology for finding the solution of the algebraic system is to apply an iterative solving method. These methods start with a guessed solution C^0 which is replaced at the algebraic system to check its accuracy while a residual error is obtained. The process continues by stepping at intermediate solutions and

try to minimise the residual error. Specifically, for an intermediate solution C^n, a condition of the following form is tested:

$$\underline{A} \cdot C^n = \underline{B} - \rho^n, \tag{8.25}$$

where ρ^n is the residual vector at iteration n. The intermediate solutions converge to an exact solution when

$$\lim_{n \to \infty} \rho^n = 0. \tag{8.26}$$

Iterative methods provide an approximate solution for the system of equations. The solver continues iterating until the residuals are smaller (or equal) to a user defined threshold (normally 10^{-6}). There are many iterative solvers that are used for obtaining approximate solutions. Amongst the most popular choices are *Preconditioned Conjugate Gradients* (PCG) and *Generalised geometric-Algebraic Multigrid* (GAMG) methods. The interested reader can find a full review of iterative methods for solving algebraic systems in the book by Ferziger and Peric [FP02].

8.2.6 Boundary and initial conditions

The boundary conditions are divided into two broad categories. These are the *Physical* and *Numerical* boundary conditions. Physical boundary conditions refer to regions of the computational domain where the fluid is coming in or out of the domain. These are known as inlets or outlets. Other boundary conditions include, but not limited to, symmetry planes, walls, boundaries where temperature is constant, walls where no heat exchange is allowed and many others. A physical boundary condition can be modelled using one or more numerical boundary conditions at the variables of interest.

Numerical boundary conditions can be classified to three categories. The first category is known as fixed value or *Dirichlet* boundary condition and requires a variable to be prescribed a constant value across the faces that compose the boundary. The second category is known as *Neumann* boundary condition and requires the gradient of the variable normal to the boundary to get a constant value. Finally, a mix of Dirichlet and Neumann boundary conditions can be applied to the boundary. This is known as *Robin* boundary conditions. Setting up numerical boundary conditions requires information regarding the location of the physical boundaries and the physical phenomena involved near them.

Initial conditions are the initial values of all the variables at the nodes of the computational mesh. These values can be defined uniformly in the domain or non-uniformly at different regions depending on the requirements of the problem. The

choice of the initial condition needs to be compatible with the physical processes that characterise the problem and in most cases wrong selection of initial conditions can lead to unstable simulations that do not converge. The selection of realistic boundary and initial conditions is critical for the success of a simulation.

8.3 Simulation of smell transport in virtual environments

The previous section gave the mathematical background of the CFD methodology for simulating odour transport in a VE. This section builds on this material and provides the simulation framework for estimating odour concentration in a series of VE case studies with different geometrical features and physical properties. The objective is to study how odour concentration changes when progressively more accurate transport simulations are applied in a VE. Numerical accuracy is adjusted using better spatial discretisation of the computational domain where the odour propagates within. This section is organised as follows. Firstly, the simulation set-up for the four VE case studies is given. This includes the description of the initial and boundary conditions and the process followed for creating realistic convection effects in these VEs. Secondly, the concentration results for different discretisation levels of the computational mesh are given. Concentration values are captured using virtual probes located at the assumed viewer's/listener's location in the VE.

8.3.1 Simulation set-up

When an odour spreads out from its container, it mixes with air and/or other VOCs creating a homogeneous gas mixture that disperses in the environment [ADP01]. The rate of odour dispersion strongly depends on the environmental conditions of the environment. In this thesis the mixture is composed of the smell of citral, a lemon like odour, and air and the objective is to estimate the concentration of citral during the transport of the mixture in the VE. Citral was preferred among other volatile compounds as it is a frequently encountered odour with known physical properties [FSQM03, Cit16].

All the CFD simulations in this thesis were implemented using the OpenFoam Application Programming Interface (API) [WTJF98] and following the *Eulerian* approach, i.e. the smell-air mixture was considered as a continuum and all the variables of interest were computed using the governing equations described in

the previous section[1]. Navier-Stokes equations (see equation 8.4) were discretised and solved along with the transport equation (see equation 8.8) for estimating the momentum of the air-odour mixture ($\mathbf{U}$), pressure (p), density (ρ), temperature (T), concentration (C) and internal energy i if the problem includes energy transfer due to temperature gradients. These variables were computed and stored at the CV centres of the mesh. As is the case with all gases, the flow transport of citral-air mixture is highly turbulent. Turbulence effects were also taken into account in the simulations presented in this chapter using a realizable $k-\epsilon$ model that offers accuracy improvements compared to the simple standard version [THWA$^+$95].

It is known that the domain's mesh fineness affects the convergence of the solution while it can increase numerical stability during the CFD simulation [SR14]. Coarse meshes yield concentration solutions with over- and undershoots because of the non-smooth transition of the odour-air mixture across the CVs. In the simulations that are described in the following four different mesh sizes were used to investigate the effect of spatial discretisation on the concentration of citral in the air. These are meshes with 1K, 10K, 100K and 1 million CVs. The coarsest mesh was refined near the surfaces and the boundaries of the VE for better accuracy while every successive refinement was uniform across the domain so as to approximately preserve the initial distribution of CVs. Mesh sizes with an order of magnitude change were selected for studying how odour concentration at probe locations changes between large spatial discretisation steps starting from very coarse up to excessively high refinement levels.

Simulation of smell transport was considered in four different VEs. The geometry models for all these scenarios are the same used for sound and light simulations and created using Autodesk Maya. The physical and geometrical characteristics of each of them are described in different subsections for better clarity. Smell propagation in every VE was simulated for 1800 sec virtual time to study how concentration evolves at the virtual sensor location for a sufficient amount of time. Initial and boundary conditions for concentration, flow rate, temperature and pressure were set up using physical values which were obtained after repeated measurements with a PID sensor, a flow meter, a digital thermometer and a digital manometer respectively. The collected numeric data were averaged over the 10 times of repeating the measurements and represent realistic estimations for the quantities of interest.

As citral is the only odour used in the experimental studies ***E2*** and ***E3***, only

[1]Another way is the *Lagrangian* approach where the odour transport is handled through tracking particles in the VE after repeatedly computing the forces being acted on them

the scenarios that are congruent with the smell of lemon were used for smell transport simulations. The full collection of scenarios is given in Figure 7.1. From these, the scenarios Bathroom, Car, Kitchen and Kitti were considered for the CFD simulations. The diffusion coefficient of citral into air was estimated using formula 8.9 for the different environmental conditions (temperature and pressure) in each of the four VEs. The virtual probe for measuring concentration values was placed at the viewer's location in the VE which is the camera position used for rendering the visual images (see Figure 7.1). The same position was also assumed for the generation of the RIR during the auralisation process.

In the following, the four scenarios are considered separately while cross-sections of the of the concentration and velocity results are given for each of them. The visualiser software used for inspecting the results is ParaView application [Aya15].

Bathroom scenario

In the bathroom scenario, smell enters the domain from the top face of a rectangular box (area = 3 cm^2) which releases odour vapour at concentration of 25 ppm. This box represents the boundary rectangle of a shampoo bottle. The temperature difference between the hot bath (40 °C) and the environment (15 °C) causes convection effects which make the smell-air mixture to be dispersed in the room at low velocity (around 3 cm/sec). In this scenario there are no outlet boundaries and the room is assumed to have no open windows/doors. Pressure was assumed to be equal to the atmospheric pressure (101.320 Pa). These temperature and pressure conditions result to a diffusion coefficient of $D_{sa} \approx 8.13 \times 10^{-5}$ cm^2/s. Smell concentration values were recorded using a virtual probe at a sampling frequency of 4 Hz. Figure 8.9 shows the smell inlet, the virtual probe and the surfaces that cause the convection effect in the environment. Figure 8.5 shows the momentum and the odour concentration profiles after 800 seconds of simulation.

Car scenario

In the car scenario, the smell source is an air freshener (area = 2.4 cm^2) releasing odour at concentration of 10.4 ppm which is dispersed in the cabin through the front air vents (area = 10 cm^2 = 0.001 m^2) that blow air at a flow rate of 4 L/min (or 0.004 m^3/min). The car's doors and windows are assumed closed. The velocity of air entering in the domain can be computed using the formula: $V = \frac{Q}{A}$, where Q is the flow rate and A the area of the inlet. This gives a velocity of 4 m/min (or around 6.7 cm/sec). Two outlet boundary conditions were assumed in this scenario,

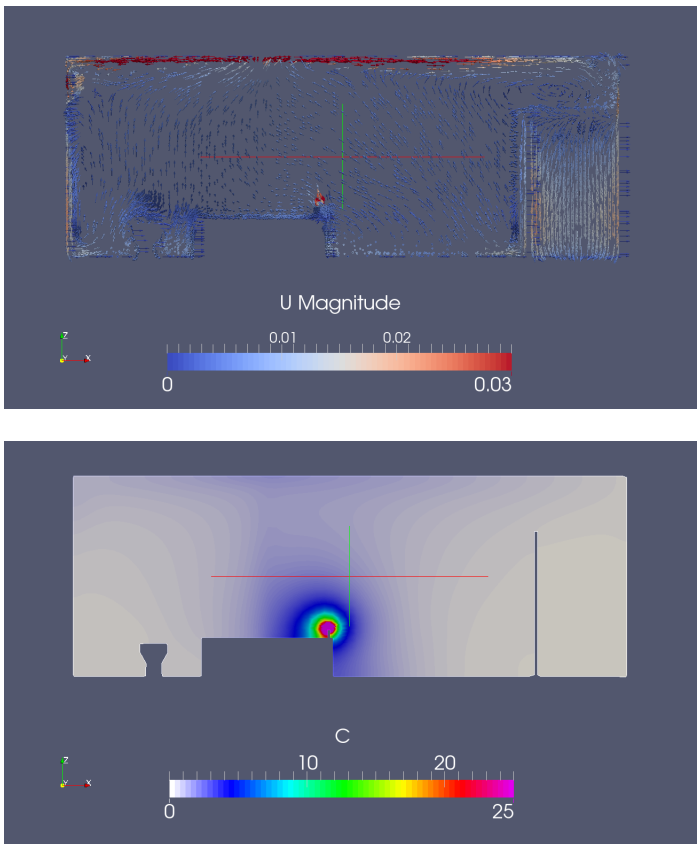


Figure 8.5: Cross-section of the Bathroom scenario. **Top**: Magnitude (m/sec) and velocity field of the mixture. **Bottom**: The odour concentration profile in the room after 800 sec of simulation.

both of them located in the car trunk. These allow the smell-air mixture to exit the domain. Temperature and pressure are assumed constant with values (15 °C) and (101.318 Pa) respectively. The temperature and pressure conditions give a diffusion coefficient of approximately $D_{sa} \approx 7.87 \times 10^{-5}$ cm^2/s. The virtual probe is located at the driver's head location and captures concentration values at a frequency of 4 Hz. Figure 8.9 depicts the boundary conditions of this scenario. Figure 8.6 shows the momentum and the odour concentration profiles after 800 seconds of simulation.

Kitchen scenario

In the Kitchen scenario, smell enters the domain from the top face of a box (area = 5.8 cm^2) at concentration of 25 ppm. This box is the boundary box of a mug that is placed next to a hot kettle. The room is assumed to have no open windows and convection effects are introduced through the temperature gradient between

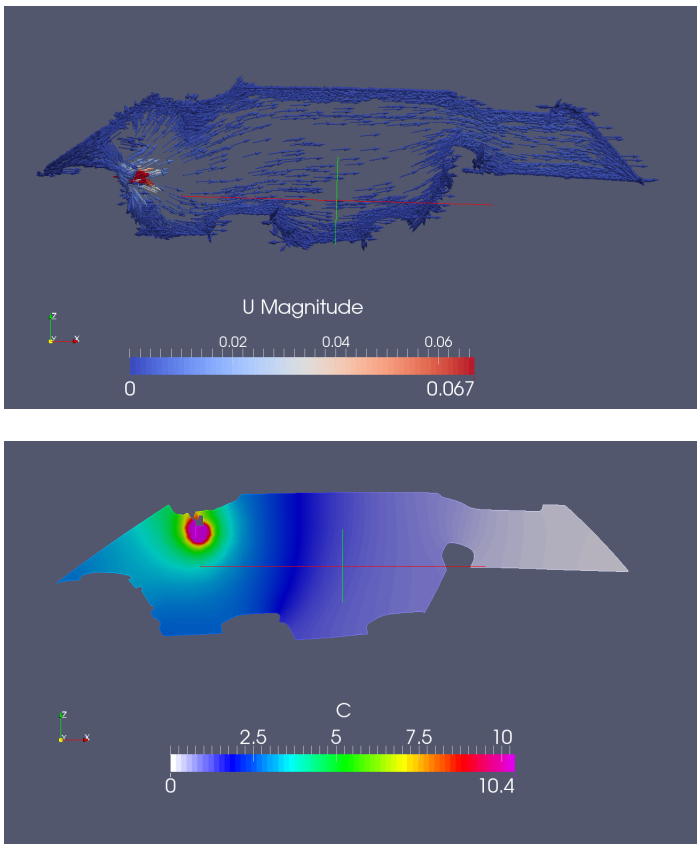


Figure 8.6: Cross-section of the Car scenario. **Top**: Magnitude (m/sec) and velocity field of the mixture. **Bottom**: The odour concentration profile in the car cabin after 800 seconds of simulation.

the hot kettle (50 °C) and the cold environment (15 °C). As the source of heat is small compared to the surrounding environment, the air-odour mixture is dispersed at a very low velocity in the room (around 0.6 cm/sec). The diffusion coefficient of citral into air for these environmental conditions is approximately 8.159×10^{-5} cm^2/s. Figure 8.9 shows a snapshot of the geometry and the boundary conditions of this scenario. Figure 8.7 shows the momentum and the odour concentration profiles after 800 seconds.

Kitti scenario

In the Kitti scenario, smell is coming out from a jar that is located on the church's main tableaux. For simplicity the jar is represented by its boundary box and the odour is coming out from the top face of this box (area = 4.8 cm^2) at concentration of 25 ppm. The VE is assumed to have three doors, the main door is behind the

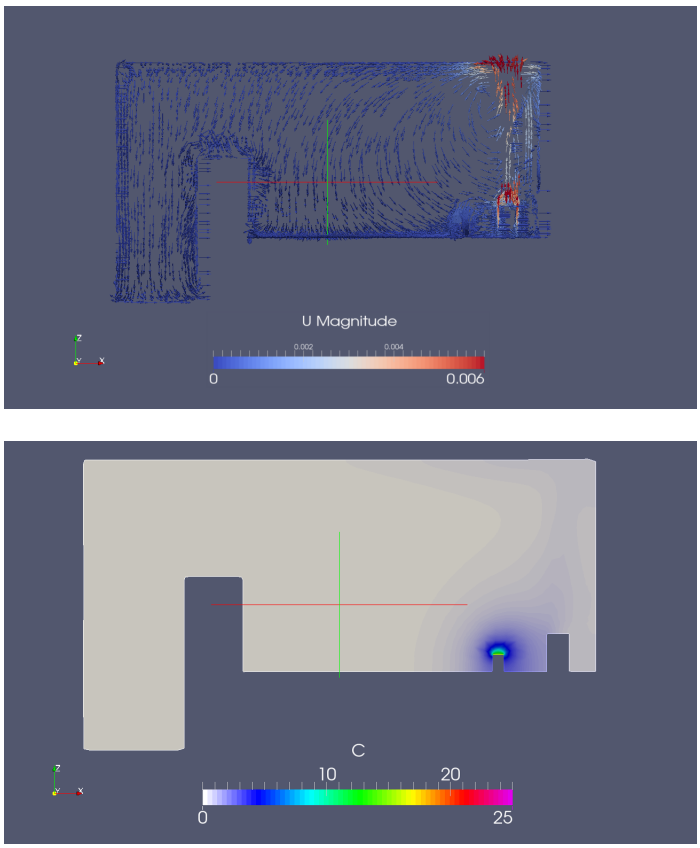


Figure 8.7: Cross-section of the Kitchen scenario. **Top**: Magnitude (m/sec) and velocity field of the mixture. **Bottom**: The odour concentration profile in the VE after 800 sec of simulation.

virtual probe and two more doors exist to the left and right walls of the smell source (see figure 8.9). All the doors are used both as inlet and outlet boundary conditions where air can enter or exit the computational domain depending on the flow direction. This set-up represents better what happens in real life where air enters or exits from the same boundaries. The air draughts enter the domain at a velocity of 16 cm/sec. The assumed temperature (18 ° C) and pressure (101.316 Pa) of the environment give a diffusion coefficient of citral into air of approximately 8.151×10^{-5} cm^2/s. The smell probe location, smell inlet patch and one of the three doors are depicted in figure 8.9. Figure 8.8 shows the momentum and the odour concentration profiles after 800 seconds.

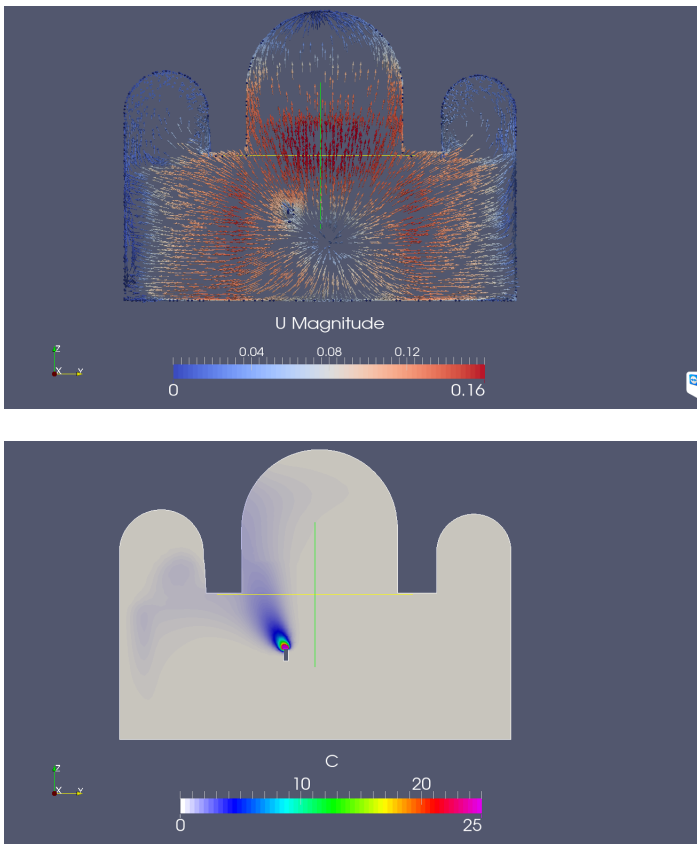


Figure 8.8: Cross-section of the Kitti scenario. **Top**: Magnitude (m/sec) and velocity field of the mixture. **Bottom**: The concentration profile in the VE (**Right**) after 800 sec of simulation.

8.3.2 Odour concentration results and computation times

Figure 8.10 shows how concentration changes over time at the virtual probe location for every scenario across the four different discretisation levels used in the simulations. Smell transport effects were simulated for a total of 1800 seconds of virtual time in each of the four considered scenarios. These graphs give a number of potentially useful results. As it can be seen, for coarse meshes the transition of the air-citral mixture near the sensor position is more abrupt therefore fluctuations are captured in the concentration function. These fluctuations are progressively eliminated as the mesh is uniformly refined. Furthermore, as it can be seen from the graphs, concentration curves are not very distant to each other. Specifically, the maximum difference between the coarsest (1K) and the finest meshes (1M) was 1.981 ppm for the Bathroom scenario, 1.623 ppm for the Car cabin scenario, 1.774 ppm for the Kitchen scenario and 1.345 ppm for the Kitti scenario. The odour

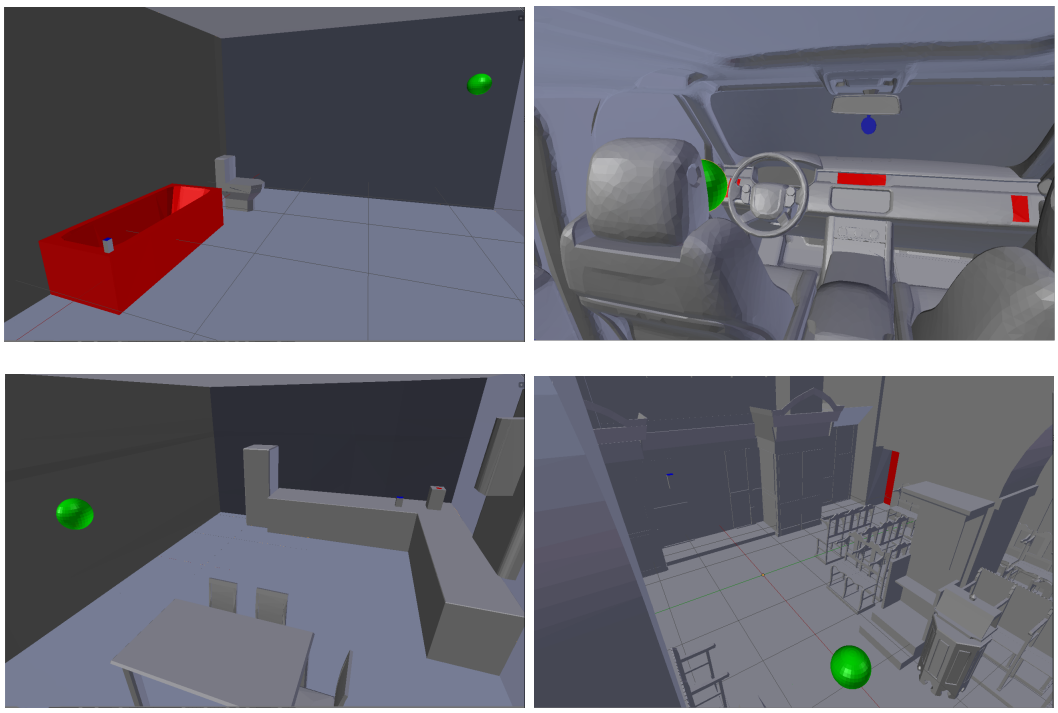


Figure 8.9: Simulation set-up of the VEs used for smell transport simulations. From **Top** to **Bottom** and **Left** to **Right**: Bathroom, Car, Kitchen, Kitti. The blue painted patches represent the odour inlets while the red patches are used for developing convention effects (inlet/outlet boundary conditions). The green coloured spheres represent the virtual probes for reading concentration values during the simulation. Rendered images of the VEs taken form the sensor's viewpoint are provided in Figure 7.1.

concentration results that are shown in figure 8.10 set the question of whether these differences are perceptually different, on average, to the users of the VEs. This question is answered in chapter 10 of this thesis by conducting a psychophysical experimental study for investigating the JND threshold for odour intensity for the smell of citral.

Another interesting point that can be concluded from the concentration graphs is that at some scenarios the curves stabilise after a virtual time moment. This is an indication that the concentration at the probe location has reached an equilibrium point where the produced and lost amounts of odour mass are approximately the same (*saturation* point). The saturation point has not been achieved for the Kitchen scenario after 1800 seconds because the smell rate of dispersion is very low in the VE. The kettle's hot surfaces do not produce sufficient heat to make the mixture circulate fast and increase the concentration at the probe location. Also,

equilibrium has not been achieved for the Kitti scenario because of the large volume of the room (560.64 m^3) and the location of the sensor. The odour-air mixture exits from the side doors of the room and only a small portion reflects back to the main room and reaches the sensor's location.

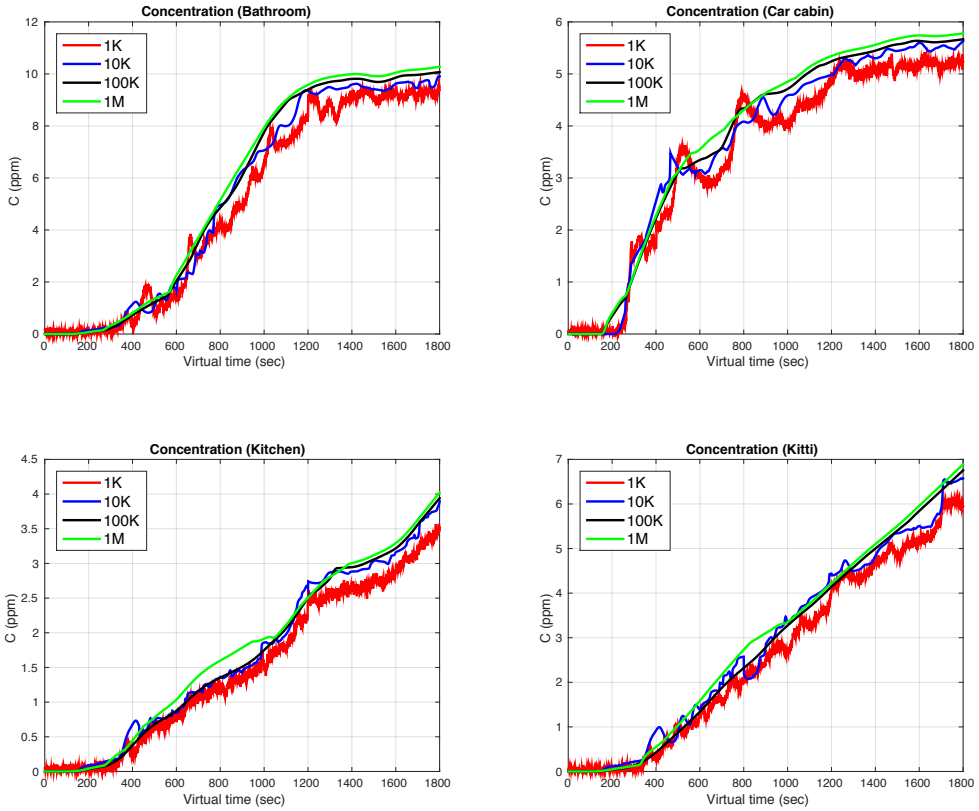


Figure 8.10: Concentration results measured for different discretisation levels (coloured curves) and across different scenarios. The concentration values were captured at the virtual probe location shown in figure 8.9.

The computation timings for every VE and mesh discretisation level are given in table 8.1. The machine used for running these simulations was an Intel Xeon E7450 at 2.40 GHz with a total of 24 physical cores and 64 Gigabytes of RAM. Domain decomposition using parallel processes over the message passage interface (MPI) was used in order to speed up the computation for every spatial discretisation level. As it can be seen the scenarios Bathroom and Kitchen give slightly more increased computation times due to the requirement for solving for internal energy i (see section 8.2.4). Heat transfer is the main reason for achieving convection effects in these scenarios and solving for the extra variable i permits

simulation of heat exchange in the VE due to temperature gradients.

	1K	10K	100K	1M
Bathroom	892.8	2583.3	7632.1	18592.6
Car cabin	695.6	2362.2	6336.7	15559.1
Kitchen	879.4	2459.9	7612.2	18588.7
Kitti	719.9	2368.5	7538.8	17990.8

Table 8.1: Physical computation times in seconds measured for the four mesh discretisation levels across the four VEs used for smell transport simulation. All the simulations ran for a total of 1800 seconds of virtual time.

8.4 Delivery of smell impulses

The previous sections concentrated in the simulation of smell transport and the estimation of physically accurate concentration values for different discretisation levels of the computational domain. This section completes the olfactory pipeline by presenting the methodology for delivering realistic smell impulses to the VE users. This methodology is based on displaying smell cues using an olfactory display that delivers physically accurate smell impulses based on the concentration values computed in the simulation stage. The full olfactory pipeline is utilised for the implementation of the experimental studies ***E2*** and ***E3*** whose methodology framework is given in chapter 4.

An olfactory display was used to deliver the smell impulses to the participants of the experimental studies. This device follows the Lorig olfactory display design (described in section 2.4.3) and is composed of several components that can be configured to be close or away from each other and/or the user depending on the scope and needs of the application. For the purpose of conducting experiments, the majority of the olfactory display parts were hidden from the participants and only the smell outlets (tubes) were visible to the subjects so as to avoid any potential bias.

For the experimental studies of this thesis, an olfactory display with two channels/outlets was used. The flow rate of each channel is adjusted using a Digital Mass Flow Controller (DMFC) as a percentage of the maximum flow rate provided which is $F_s = 1000$ ml/min. One of the channels was used to release smell stimuli (citral) and the other as a control channel for releasing clean air and removing residual smells from previous trials during the experimental procedure. Figure 8.11 shows a set-up of the olfactory display during a testing session for measuring its

output concentration and flow rate. Testing sessions were used to calibrate the device and ensuring that physically accurate smell stimuli are to be delivered during the experimental sessions. During the experimental trials, the device was in compact form under the desk and only the two outlet tubes were visible and located in places near the user's nose (see Figure 10.2).

As can be seen from figure 8.11, an air compressor is used to supply a continuous air stream of pressure 2 bars to the olfactory display. The stream is filtered through an in-build air filter and directed to the two DMFCs. Each of the DMFCs contains a computer controlled diaphragm that allows any percentage of the incoming stream to pass. The modulated air is aromatised by passing through a reservoir where citral in liquid form is stored and released at the edge of the outlet tube. Lastly, a PID detector (see section 2.4.4) is used to record output concentration values. These values are directed to a Raspberry Pi single board computer which outputs them in real time to a computer monitor for inspection and processing. The set-up and calibration of this device was work in common with two other members of the visualisation laboratory.

Based on the maximum flow rate of a DMFC ($F_s = 1000$) ml/min, The output volume of a smell burst can be calculated in ml using the formula:

$$V = F_s S_p t, \tag{8.27}$$

where S_p is the set point and t is the time of the smell burst required in minutes. For example, for an $S_p = 0.5F_s$ and burst time $t = \frac{1}{60}$ min the output bursts have volume of 8.3 ml. After a specific volume of air is released by the DMFC, it passes through a sealed reservoir where it is enriched with the smell which is stored in liquid form. Finally, the burst is delivered to the user using flexible solenoids (of approximately 1 m length each). Pressure measurements were conducted at the joints of the device during the test sessions to ensure that the pressure level remains invariant throughout the device use.

The operating range for guaranteed concentration accuracy for the citral VOC is $I = [C_0, C_{\max}] = [1.2, 11.2]$ ppm for a burst time of 1 sec. These values were obtained by collecting concentration data with a PID sensor at different output volumes of the air-citral mixture (see formula 8.27 and figure 8.11). The collected data were fitted with the linear prediction model $C = 0.613V + 1$ which gives the above range of concentration values for the volume range $[V_0, V_{\max}] = [0.3262, 16.659]$ ml of the output smell bursts. This model has coefficient of determination $R^2 = 0.9919$ which indicates almost perfect fit to the collected data.

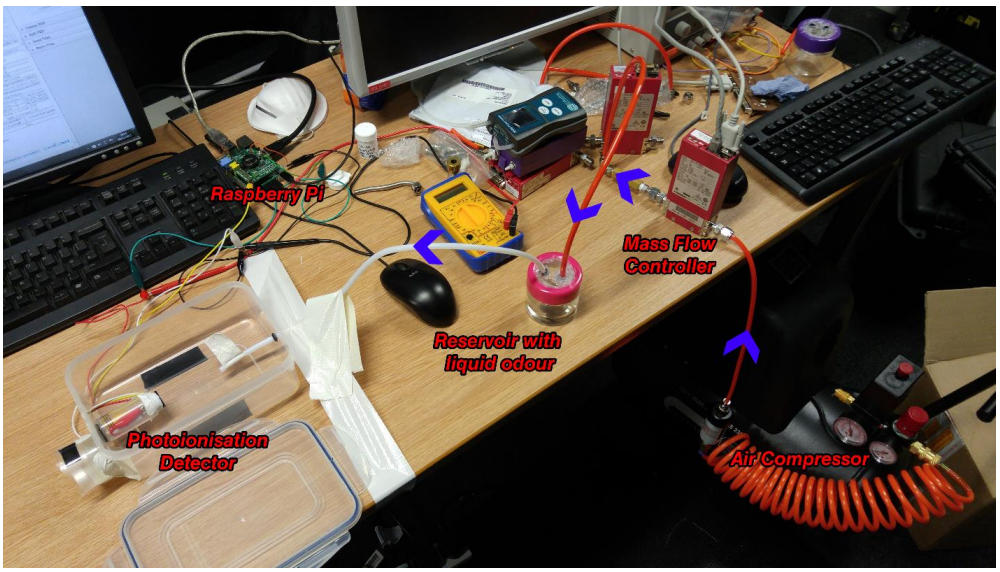


Figure 8.11: The olfactory display used in this thesis. The blue arrows show the air flow direction. Starting from the air compressor to the mass flow controllers and after that through the smell reservoir the air stream is aromatised and delivered to a PID for inspection (during testing) or the user (during an experimental trial).

The onset and rise times of the olfactory display were also measured using the PID sensor and found 240 ms and 73 ms respectively. These values show that the delivery of any other sensory stimulus needs to be delayed with at least 240 ms delay if synchronous delivery of all the sensory stimuli is required in the application. The onset and rise times of the olfactory display were used for the implementation of the experimental study ***E3*** were synchronous delivery of visual, auditory and olfactory cues is required.

8.5 Discussion

PID sensors (see section 2.4.4) have certain limitations on the number of VOCs that are able to be detected. Manufacturers design PIDs with the objective to measure concentration values of frequently encountered VOCs or those VOCs which are harmful when inhaled in high concentrations. Therefore, only a small portion of the trilion odorous VOCs [BMVK14] can be detected by these sensors. The detection of a VOC by the PID sensor allows to estimate its concentration at any given point in the environment. For delivering accurate smell stimuli using olfactory displays, it is necessary to measure the range of concentration values that can be delivered for a given VOC thus, it is important to ensure that the output bursts

can be detected by the PID sensor. As any PID is able to detect only a subset of all possible VOCs, a limitation is set on the number of different smells that can be delivered in VR applications.

As mentioned in section 3.1.3 many smells are composed of multiple VOCs that have no odour as individual entities or their odour is not known. However, they can trigger a commercial PID. Detecting only some VOC components of a smell stimulus and measuring their concentration does not necessarily mean that the same concentration range is valid for the original smell that is composed of multiple VOCs. This is a limitation as the vast majority of smells is composed of many VOCs of different properties. Future work needs to take into account GC-MS analyses (see section 2.4.4) for investigating the proportion of every VOC in a given smell and later use this information for estimating a concentration range based on these proportions.

The smell of citral is an odorous VOC that is composed by only one component (*single molecule* VOC) and can be detected by the available PIDs bought for conducting this research thus it overcomes the limitations noted previously. Although other detectable VOCs could be used instead of citral, for example, pinene (smell of pine tree) or banana oil their congruency with the existing scenarios was limited. Citral is congruent with four of the scenarios used in this thesis and this is an additional reason for using it.

8.6 Summary

This chapter described the methodology followed for the estimation and delivery of olfactory cues. The material of this chapter is the basis for the experimental studies ***E2*** and ***E3*** of this thesis. Firstly, the CFD mathematical framework for estimating odour concentration in a computational domain is given. This estimation is achieved through simulation of the air-odour plume in a VE considering both convection and diffusion physical effects. The simulations were implemented in a range of different VEs with different geometrical features and physical requirements. Concentration values were recorded at a virtual probe location which represents the user's location in the VE and was also used for physically based simulation of graphics and acoustics. Secondly, the delivery of physically accurate concentration values to the users using an olfactory display was discussed. The olfactory display was tested for validating its operating range in terms of output concentration an flow rate accuracy.

Chapter 9

Just noticeable difference threshold for perceptually assessing smell simulations in virtual environments

9.1 Introduction

Multi-sensory VEs can provide high levels of fidelity by simulating multiple, physically accurate sensory stimuli [CDB09]. The introduction of realistic olfactory stimuli can significantly increase the sense of presence and provide high levels of immersion to the user [Che06]. The sense of smell strongly affects how people react, feel and perform actions in real life situations. There is evidence that humans recall memories associated with previously encountered scents [HE96]. The technical challenge of estimating physically accurate olfactory cues was addressed in the previous chapter through the application of fluid dynamics for simulating smell transport in VEs (see chapter 8). Estimated concentration values are then directed to an olfactory display for delivery to the users at correct intensity and flow rate.

Many limitations of the human sensory system have been exploited in the past in order to reduce the computational complexity of rendering systems with little or no perceived quality difference at the delivered visual or auditory stimuli. The Human Olfactory System (HOS) is not an exception to this rule as it is not able to perceive quality improvements that exceed certain perceptual thresholds. Numerically accurate smell transport simulations that yield stimuli that exceed such a threshold are considered wasteful for human requirements as the average user is not

able to discriminate any quality improvement; therefore the computational resources can be saved and used elsewhere.

The work presented in this chapter examines how the Just Noticeable Difference (JND) threshold for odour intensity changes over a wide range of concentration values provided by an available olfactory display. The JND threshold is estimated in an experimental study via pairwise comparisons of smell stimuli delivered to the participants using a custom GUI. In the context of this chapter, the JND threshold expresses the minimum concentration difference that humans can consciously perceive two smell impulses as different in terms of their intensity.

The JND threshold can be used to infer whether numerically accurate smell transport simulations of progressively finer spatial discretisation level can yield odour intensities that are perceptually distinguishable to the average user. The effect of better mesh fineness on the concentration values recorded in a virtual sensor located in the viewer's/listener's location was presented in section 8.3.2 of chapter 8 for a series of VE case studies. These results are combined with the JND threshold estimations presented in this chapter to conclude whether better numerical accuracy is a strict requirement in smell transport simulation problems.

The hypothesis made at the beginning of this experimental study postulates that the JND threshold is quite large and the differences between progressively accurate smell transport simulations is lower than the estimated threshold. The validity of this assumption would mean that computational resources could be saved by avoiding numerically accurate simulations that provide no perceptual gain to the user.

This chapter is organised in the following sections. Firstly, related work for estimating the JND threshold for odour intensity is presented (section 9.2). Secondly, a psychophysics experimental framework for estimating consecutive JND thresholds for odour intensity in the range of concentration values given by an olfactory display is discussed (section 9.3). Thirdly, the results of the psychophysics experiment are given in section 9.4. Finally, the last section discusses limitations and possible extensions of the current work (section 9.5).

9.2 Related work

JND thresholds have been extensively used in the past both in the visual and auditory domains. In vision, JND thresholds are used at the core of the Visible Differences Predictor (VDP) computational model [NMDSLC15] along with contrast sensitivity functions to encode the non-linear response of the eye photoreceptors when

they are triggered by light. In audio, JND thresholds for frequency and loudness have been extensively used in the study of auditory masking which is the cornerstone of many audio compression schemas [Gel09, JJS93]. Considering olfaction, experimental studies on patients make use of JND thresholds for the detection of diseases related to smell deficiencies [HCG+11]. JND thresholds have also been used for researching the possibility of developing incidental learning or memory abilities for different olfactory or gustatory cues [KPK04].

Perceptual thresholds for odours and specifically for the smell of citral which is used in this experimental study, have been estimated in previous work but have concentrated mostly in estimating the detection threshold of a smell and not the intensity JND threshold [SW55, MC80]. Moreover, the olfactory display technology used in these studies did not offer the same advantages as the modern, temporally accurate, devices used today. The experimental methodology has also progressed and new methods are used for accurately predicting perceptual thresholds [KP09].

9.3 JND experimental layout

The psychophysics experiment for estimating the JND threshold for odour intensity was designed using a custom GUI where participant's task was to select the most intense smell stimulus among two olfactory stimuli presented in different temporal order. This section describes the experimental framework for estimating smell intensity JND thresholds in a range of concentration magnitudes. Details about the materials used, the procedure and the subjects that took part in the experimental study are also given in this section. The JND threshold estimation allows to assess whether the recorded concentration curves for different mesh discretisation levels are perceptually equivalent (see section 8.3.2).

9.3.1 Experimental design

The experimental methodology used for estimating the JND threshold was the two Interval Forced Choice (2IFC) method [KP09]. At every trial two olfactory stimuli were presented with an Interstimulus Time Interval (ITI) between them. The first is a constant concentration stimulus, known as *pedestal*, while the second is known as the *varying* stimulus because it can take different values at different trials. The varying stimulus can get a number of concentration values from a predefined set of olfactory stimuli that are compared with the pedestal stimulus. Ten trials were performed for each comparison pair (pedestal, varying) for the different values of the varying stimulus. The presentation order of the two smell impulses (pedestal-

varying or varying-pedestal) was also randomised across the experimental trials to avoid any potential ordering bias. The odour selected for the implementation of this experiment was the smell of citral. Citral is a lemon like smell, frequently found in everyday life, thus, the majority of the participants could easily detect it.

The JND threshold is the magnitude that needs to be added to the pedestal so as the resulting olfactory stimulation has magnitude that is perceptually more intense than the pedestal in 75% of the total trials [KP09]. During each trial, the two stimuli were presented with an ITI interval of 4 seconds while the Stimulus Exposure Duration (SED) was 1 second for each of the two delivered bursts. A short SED was selected to avoid user adaptation to the released smell while the ITI was sufficiently long to permit the participants to rest before inhaling the second stimulus. Figure 9.1 shows pictorially the temporal order of stages in an experimental trial.

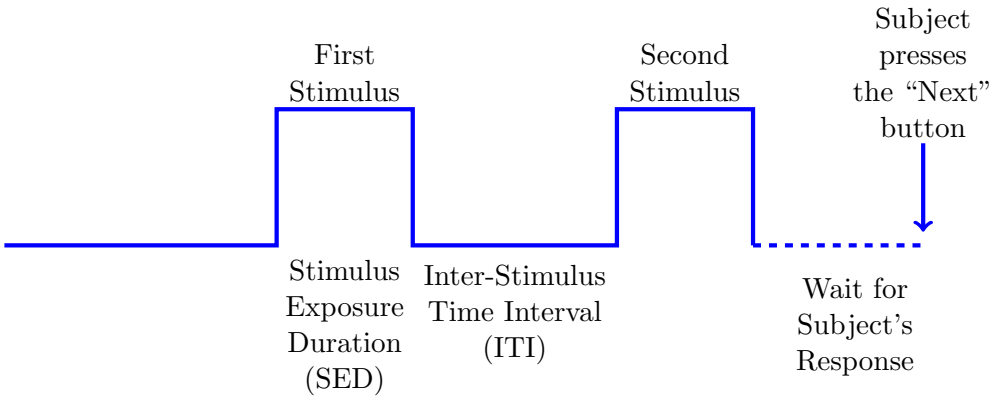


Figure 9.1: Stages of an experimental trial during the experiment ***E1***. The figure depicts the Interstimulus Time Interval (**ITI**) between two successive stimuli and the Stimulus Exposure Duration (**SED**) of a delivered olfactory stimulus.

For this experimental study, the same olfactory display hardware that described in section 8.4 was used. As explained, this device guarantees concentration accuracy of the delivered stimulus when is operated in the range $I = [C_0, C_{\max}] = [1.2, 11.2]$ ppm. These guidelines were followed in this study and all the presented olfactory stimuli had concentration from the range I. All the participants were able to detect the lemon like odour when they sniffed C_0 stimulus during a training stage before the main experimental session. This means that the whole range of stimuli contained in I were higher than the absolute detection threshold. The JND estimation would not be possible if the delivered stimulus could not be detected by the participants.

As there was no previous knowledge for the existence of any JND threshold in the range I, starting with C_0 as the pedestal, pairwise comparisons of C_0 with a number of varying stimuli that uniformly discretise I were planned until the average performance among all the subjects would reach a proportion of 75%. For a step s the interval I was subdivided to three mutually exclusive subsets:

$$\begin{aligned} I_1 &= [1.2 + s, 1.2 + 2s, \cdots, 1.2 + 8s] \\ I_2 &= [4.8 + s, 4.8 + 2s, \cdots, 4.8 + 8s] \\ I_3 &= [8 + s, 8 + 2s, \cdots, 8 + 9s]. \end{aligned} \tag{9.1}$$

A step size s equal to 0.4 was selected so as to keep a reasonable number of varying stimuli for each of the three subintervals and be able to span the whole range in the extreme case that no perceptual different magnitude existed between C_0 and any of the stimuli in I_1, I_2, I_3.

Starting with comparisons of the pedestal C_0 with the stimuli in interval I_1, if there was no stimulus that triggers a perceptual difference, denoted as C_1 the experimentation would be carrying on in a subsequent experimental session for the same pedestal and the stimuli with magnitudes in the interval I_2 until the first stimulus that triggers a perceptual difference. If there was no such stimulus, the pedestal would be compared with stimuli in I_3 in a subsequent experimental session. In case C_1 could be found, there is no need to continue experimenting for the rest of the stimuli in any of the I_2 and/or I_3 as it was necessary to find perceptual differences starting from C_1 as the new pedestal and estimate a new JND odour intensity threshold.

The first time that a JND threshold was found a new experimental session was initiated by adjusting the step for spanning the rest of the interval I using Weber's law. This law is applied in the olfactory domain [AG98] and in this context can be written as follows:

$$\frac{\text{JND}}{C_p} = \frac{C_f - C_p}{C_p} = k, \tag{9.2}$$

where C_p is the pedestal and C_f the final intensity stimulus that triggers a perceptual difference. The constant k can be obtained from the first JND, denoted as JND_1, and the first pedestal $C_0 = 1.2$ ppm. Using the constant k, we can obtain an estimation of the final stimulus that elicits perceptual difference starting from C_1 as the new pedestal and applying again Weber's law. In case the estimation is not in the interval I, no further experimentation is required as it cannot be delivered by the available olfactory display. If not, a subinterval with new discretisation step is derived and

a new experimental session is initiated. The new subinterval includes Weber's law estimation as the potential stimulus that elicits perceptual difference. The new step is computed so as the new interval does not exceed 10 varying concentration magnitudes. Experimental sessions with more than 10 varying stimuli can be tedious for the participants while this also effects the quality of the collected data.

This adaptive method allowed for the speed up of the process of finding a new stimulus that triggers perceptual difference without comparing intermediate pairs that would be perceptually similar using the initial small step of 0.4 ppm. The derivation of a new step enabled avoiding unnecessary experimental trials. Using this methodology, two JND thresholds were found in the operating range I. The first one was included in I_1 interval while the second one was found after one step adjustment using an interval of 10 varying stimuli. These findings are thoroughly discussed in the results section (see 9.4).

9.3.2 Participants

The estimation of the JND threshold for odour intensity was composed of two experimental phases with 10 participants each for a total of 20 people. All the participants were assigned randomly to one of the two phases before starting the experiment. From them, 11 were men and 9 were women with ages between 19 and 53 years old and from various academic disciplines. The average age of all the participants was 31.6 years old. No participant reported olfactory deficits or any kind of temporary or permanent anosmia problem. The sample of participants was opportunistic as it was comprised by university students and academic staff. All the subjects participated voluntarily in the study and received no payment.

9.3.3 Materials

For this study a 15.6″ laptop computer was used for controlling the olfactory display and the GUI. A snapshot of the GUI used is given in Figure 9.2. Two tubes arrived near the nose from the olfactometer, one that delivered smell bursts at every experimental trial and one air control channel that was evacuating the old smell stimuli by releasing humidified air at an exact flow rate of 1000 mL/min during the ITI. The experiments were conducted under normal indoor lighting conditions in a room where air was constantly ventilated. The duration of each experimental phase was two days with a break of one day before the start of the second phase for re-calibrating the olfactory display. Figure 9.3 provides a schematic experimental setup followed in this study.

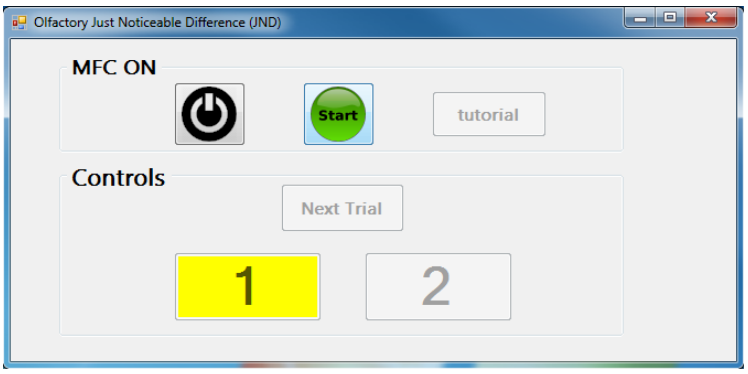


Figure 9.2: Snapshot of the software used for the ***E2*** experimental study. It shows an instance of the GUI when the first smell impulse is delivered. After the end of each trial the participant selects the most intense smell impulse by pressing the button labeled **"1"** or **"2"** respectively.

The distance of the nose away from the smell delivery point was kept constant at 7 cm using a chin rest located in front of the laptop. An air compressor located under the table was used to provide two air channels for the experiment at a constant pressure of 2 Bars. The first channel was aromatised by passing through a smell reservoir where citral in liquid form was stored. The maximum delivery flow rate for both the air and smell streams was 1 L/min. This flow rate was not high enough to cause tactile stimuli at participant's face at any given concentration in the range I. Both the aromatised and the control channels were controlled using the olfactory display's mass flow controllers which were connected through USB ports on the laptop.

The odour of citral was used as the delivered olfactory stimulus in this experimental study. This is a frequently encountered volatile organic compound which has a low detection threshold that ranges between 25 to 350 ppb [Cit16]. All the participants were able to detect it while no participant reported the smell as unpleasant. The range I containted intensity stimulations that varied from faint up to more intense smell impulses without being irritating for the people who sniffed them.

9.3.4 Procedure

The experimental procedure was initiated with a tutorial where the experimenter could explain the scope of the study and the task required from the participants. Every participant had to smell the two bursts and then decide which one was the most intense by pressing one of the GUI buttons provided. The total length of

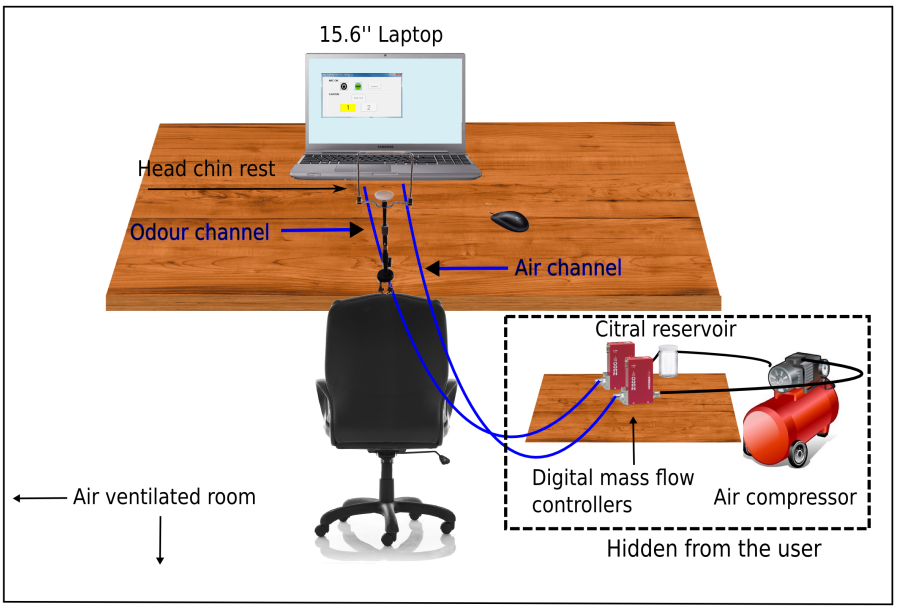


Figure 9.3: Hardware setup for the ***E2*** experimental study.

the experiment was approximately 25 minutes. During the tutorial session, pairs of (C_0, C_0) stimuli were presented and the participant was asked if the smell was familiar or not. All the 20 participants were able to recognise the delivered odour with no problem.

In the main experimental session, three seconds before the first smell release, the participant was prepared by a visible countdown above the disabled button "**1**" while during the first burst, button "**1**" was being highlighted yellow to signify the release of the first stimulus (see also Figure 9.2). After the first smell impulse, button "**1**" was no longer being highlighted and the ITI was taking place followed by a three second visible countdown above button "**2**". During the release of the second burst, button "**2**" was being highlighted yellow to signify the release of the second stimulus. After the end of the second impulse, button "**2**" was no longer being highlighted and both buttons were being enabled so as the user could select the most intense stimulus. The purpose of the countdown was to prepare the participants to adjust their breathing and avoid exhaling during the burst time as much as possible.

The participants could go to the next trial by pressing the **"Next Trial"** button. The purpose of this control button was twofold. Firstly, to allow subjects to rest before going to the next trial and secondly, to reduce the possibility of subject's adaptation to the presented odour by releasing odour bursts only when was absolutely necessary. Appendix D includes ethical considerations for this and

the other studies of this thesis.

9.4 Results

In this section the results of the JND experimental study are presented. Two JND thresholds for odour intensity were found in the operating range I provided by the olfactory display at two different experimentation phases. The first estimated stimulus that elicited a perceptual difference had magnitude in the subinterval I_1 while the second was obtained using one step adjustment with the adaptive method described in the design section (see section 9.3.1).

Phase 1 started with a pedestal stimulus $C_0 = 1.2$ ppm while there were 8 varying stimuli to compare it with (see interval I_1 at Equation 9.1). Figure 9.4 depicts the proportion of correct answers across all 10 subjects participated in this phase and the logistic Psychometric Function (PF) fitted to the data. The logistic function is written as:

$$F_L(x; C, S) = \frac{1}{1 + \exp(-S(x - C))}, \tag{9.3}$$

where C is the parameter that corresponds to the magnitude that elicits a perceptual difference starting from the C_0 pedestal stimulus and S the slope of the psychometric function. The PF is used to estimate both of these parameters. The logistic PF was preferred over others as it most accurately fitted the obtained data using deviance criterion as a model comparison metric [KP09]. Model comparisons using deviance criterion are discussed later in this section. For the 2IFC method, the stimulus that corresponds to 75% of correct answers is given through the PF and corresponds to the magnitude $C_1 = 3.4996$ ppm. Therefore, the first JND threshold is $\text{JND}_1 = C_1 - C_0 = 2.2996$ ppm.

The slope of the PF expresses how subjects' performance changes with increasing concentration magnitudes. The estimate for this value was $S_1 = 1.4214$ ppm. The standard errors for the PF estimates for C_1 and S_1 were obtained based on a parametric bootstrap analysis. Bootstrap analysis is a technique used to estimate the error of the parameter estimations without repeating the experiment multiple times and compute the standard errors of obtained estimations. Starting with a given set of data, bootstrap performs resampling and generates a number of new datasets that have the same distribution as the input dataset. In this study $N = 1000$ data sets were generated using the collected data of phase 1. For all the generated data collections the parameters C_1 and S_1 were estimated and their standard deviation was computed. These were $\text{SD}_{C_1} = 0.1154$ ppm and $\text{SD}_{S_1} = 0.2495$

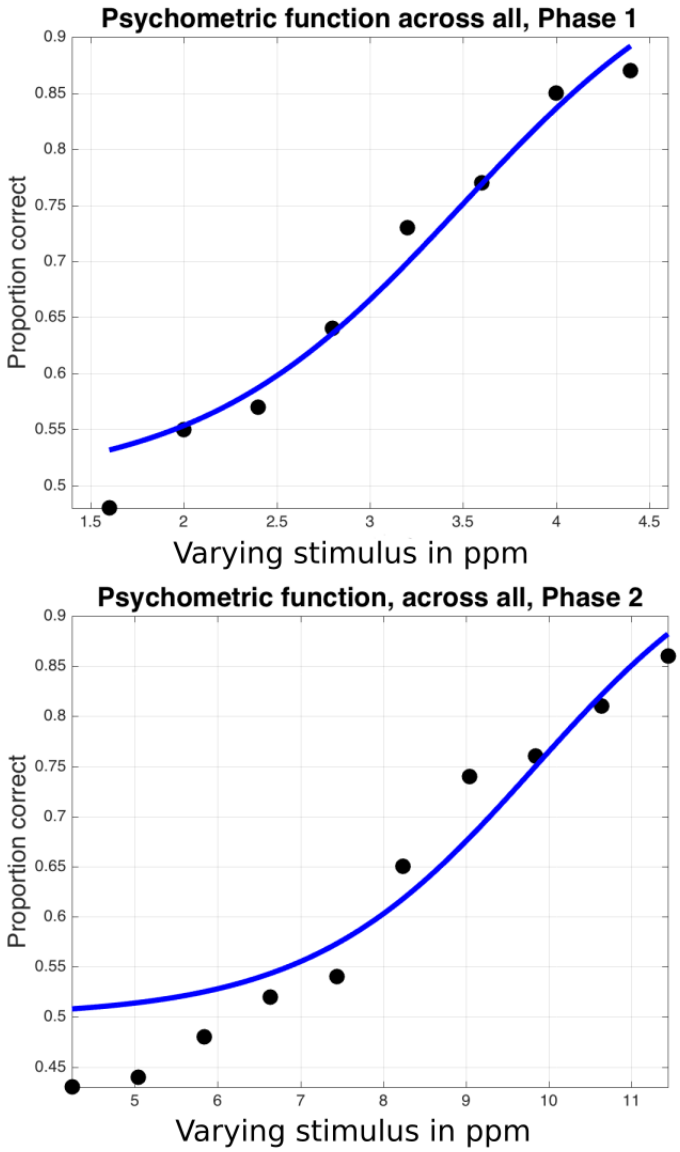


Figure 9.4: Proportion of correct answers across all subjects in Phase 1 (**Top**) and Phase 2 (**Bottom**) as a function of the varying stimulus (ppm). The blue coloured curve is the logistic PF fitted to the data in both experimental phases.

ppm.

The goodness of fit for the selected PF was also computed to check if the selected PF sufficiently fits the data. Deviance is frequently used in hypothesis testing between statistical models to measure model's goodness of fit to given data depending on model's predictions for the unknown parameters [NW72]. In this case the parameters C, S were estimated using the experimental data. The p-value for

deviance found $P_d = 0.6940 > 0.05$ meaning that the Logistic (**L**) PF is a representative statistical model for the given data. The same criterion with other tested psychometric functions such as Weibul (**W**), Gumbel (**G**), Cumulative Normal (**CN**) and Hyperbolic Secant (**HS**) PFs resulted in a lower deviance p-value meaning that these models do not sufficiently describe the obtained data. The deviance values for all the tested models are given in Table 9.4.

The Weber constant for the results of phase 1 was $k = 1.916$. Using the C_1 as the new pedestal, Weber's law (see formula 9.2) estimates that the new stimulus that elicits a perceptual difference is approximately 10.20 ppm. Using this outcome the step was adjusted as $s = 0.8$ ppm while a new experimental phase with a set of 10 varying stimuli was initiated. This set contained the stimuli with concentrations:

$$I_a = \{C_1 + s, C_1 + 2s, \cdots, 11.2\} \tag{9.4}$$

The last varying stimulus was 11.4996 ppm which is out of the olfactory display's operating range, therefore it was replaced with interval's I maximum. The proportion of correct answers and the PF function are given in Figure 9.4. The stimulus that elicited a second perceptual difference was found to be $C_2 = 10.132$ ppm meaning that the second JND threshold is $\text{JND}_2 = C_2 - C_1 = 6.6324$ ppm. This estimation is relatively close to Weber's law prediction and the small difference is attributed due to experimentation error. The slope of the new PF is $S_2 = 0.77$ ppm while the standard deviations for both estimations are $\text{SD}_{C_2} = 0.23$ ppm and $\text{SD}_{S_2} = 0.13$ ppm using the same number of generated data sets as in phase 1 for the parametric bootstrap. The deviance p-value for assessing goodness of fit was $P_d = 0.34 > 0.05$, higher than all the other PFs tested (see Table 9.4).

	L	**W**	**G**	**CN**	**HS**
Phase 1	0.6949	0.581	0.579	0.631	0.629
Phase 2	0.34	0.301	0.293	0.318	0.325

Table 9.1: Deviance p-values for the different statistical models tested for goodness of fit for each of the two experimental phases. The larger the deviance value the better the prediction model describes the obtained data.

9.5 Discussion

The smell transport simulation results for different spatial discretisation levels presented in section 8.3.2 of the previous chapter and the JND estimations for odour

intensity presented in this chapter yielded a number of potentially interesting findings. The experimental results clearly revealed that the human olfactory system is relatively insensitive at small intensity variations of the delivered smell stimulus of citral. This result is confirmed from the high JND thresholds found at the whole operating range provided by the available hardware. The low standard deviations SD_{C_1} and SD_{C_2} of the two stimuli estimates C_1 and C_2 that elicited perceptual differences show that the majority of the subjects had difficulty in perceiving small intensity differences between the pedestal and the varying stimuli. For olfactory stimulations of higher concentration the JND threshold is higher meaning that intensity variations are much more difficult to get perceived.

Although fine spatial discretisation levels of the computational domain can provide higher numerical accuracy in many applications, for olfaction, the JND threshold is clearly higher than the difference in odour intensity between the two extreme mesh sizes for the scenarios tested. This confirms the hypothesis that accurate spatial discretisation of the computational domain for the four scenarios presented in section 8.3 does not elicit quality improvements that can be consciously perceived, on average.

Although citral is an easy to detect odour, its JND threshold was quite high. This result enabled to hypothesise that odours that are hard to detect might generally give higher JND thresholds. The validity of this assumption, which is worth checking in future work, would mean that the results of this study would hold for these odours as well.

9.6 Summary

This work proposed in this chapter estimates the JND threshold for odour intensity over a wide range of concentration values. The experimental methodology is based on pairwise comparisons of olfactory stimuli which are presented to the user at different temporal order using an olfactory display, a specialised hardware for delivering correct smell intensity smell impulses. The results of this psychophysics experiment show that people are not generally able to detect small concentration variations ($JND_1 = 2.2996$ ppm.). This outcome is even more prominent at higher concentration levels where the JND threshold for the intensity of the citral odour is much higher ($JND_2 = 6.6324$ ppm).

These findings combined with the smell transport results of the previous chapter suggest that spatial discretisation refinements do not offer any advantages in the perceived smell intensity experienced by the users of the VEs tested. This

outcome means that computational resources could be saved and allocated to the computation of other, possibly more demanding, sensory stimuli provided in a multisensory VE framework such as visuals. The experimental methodology proposed in this chapter can be used as a guideline for estimating JND thresholds for other smells or combinations of multiple odorous components depending on the available olfactory display.

Chapter 10

Resource allocation for tri-modal virtual environments

10.1 Introduction

The work presented in this chapter extends the experimental methodology for resource allocation from bi-modal (see chapter 7) to tri-modal VEs where the sense of olfaction is considered along with the senses of sight and hearing. The objective is to provide a load balancing framework where diverse sensory stimuli are delivered to the user. For this purpose an experiment is conducted to capture people's quality preferences when distributing a computational budget to the three senses and across a variety of scenarios. The collected data are subsequently utilised in the construction of a prediction model capable of estimating not only the percentage of the total budget that needs to be devoted to improve the visual and acoustic stimulus quality but also to infer whether olfactory impulses should be delivered to the user depending on the available budget size and scenario.

The findings of chapter 9 indicate that numerical accuracy of smell transport simulations via finer spatial discretisation has no perceptual benefit to the average user. The JND threshold for citral intensity is higher than the difference between any pair of concentration curves at any virtual time point (see figure 8.10). In addition to these results, previous evidence suggests that high odour intensity is considered unpleasant regardless of the type of the delivered smell [JCT95], therefore, higher smell intensity does not necessarily mean better quality.

The concept of "smell quality" is not well established in the literature for two primary reasons. Firstly, the way the olfactory reception mechanism operates is not fully known and interpretations to many perceptual characteristics of the

HOS are still not well established [HCL03, Nic09]. Secondly, evidence shows that humans tend to evaluate olfactory cues in a subjective manner depending on previous experiences with a perceived smell [SP04, Nef98], therefore, quality preferences are biased from these experiences. For these reasons, the experimental layout that is presented in this chapter proposes that smell cues will be delivered using a two level scale that defines smell ON or OFF during the resource allocation task.

In addition to the interlinked sliders for adjusting the audio-visual quality, the experimental procedure requires the users to use two more controls for turning the smell stimulus ON or OFF respectively. A new custom GUI that contains the new controls is used for conducting this experimental study. When the smell is set to OFF, the total budget is completely allocated between the costs for a visual and an auditory stimulus. These costs correspond to quality levels that are delivered to the user. In case the smell is set to ON, a percentage of the total budget is immediately reserved for the smell stimulus and the rest of the budget is available for audio-visual quality improvements as previously. The GUI design makes the user aware of the budget amount that is left for audio-visual quality improvements after reserving the portion required for the smell cues.

The experiment was implemented with five distinct budgets across three different scenarios. One more scenario was used for training before the main experimental session. The experimental design is within participants and each participant was requested to judge the best perceived quality for all the 15 budget-scenario combinations. The presentation order was randomised to avoid any potential bias.

This chapter is structured as follows: firstly, section 10.2, explains how the methodology for audio-visual cost interactions in bi-modal VEs can be extended to include the theoretical cost for olfactory stimuli. Secondly, the experimental layout is given in section 10.3 and includes design, materials, participant information and experimental procedure. Thirdly, the experimental results are discussed in section 10.4. Fourthly, a statistical model is built and validated using a second experimental study in section 10.5. Lastly, limitations and possible future extensions are discussed in section 10.7.

10.2 Visual-auditory-olfactory cost interactions

The experimental study ***E3*** makes use of the normalised cost functions for visual and auditory quality stimuli given by formulae 5.9 and 6.6 for the senses of vision and hearing respectively. As explained in the resource allocation framework for bi-modal VEs (see section 7.2), normalised cost estimations allow to investigate which

of the two senses is more important for a given budget size without considering the problem of large differences in physical computation time for rendering an image and an impulse response. Also, normalisation of the costs makes the experimental results independent of the algorithmic strategy selected for computing the audio and visual stimuli.

In order to be able to investigate the importance of each of the three senses in a tri-modal experimental set-up, it is necessary to define a theoretical cost associated with the release of a smell impulse when the corresponding GUI control is enabled by the user. This metric needs to be consistent with the normalised cost functions defined for the visual and auditory quality levels. When a smell is released, the theoretical cost is automatically subtracted from the total budget to indicate user's preference to experience the olfactory stimulus.

The theoretical cost for an olfactory cue is estimated using the concentration results over virtual time presented in section 8.3.2. Specifically, when the control for smell is enabled, a smell impulse of 1 second and concentration of $C_1 = 3.4996$ ppm is released. The selection of the stimulus C_1 is explained later in this section. For a given scenario, denoted by S, and a mesh size, denoted by M, the notation VT_M^S denotes the virtual time moment when the simulation reaches the concentration C_1 at the probe location for the chosen mesh and scenario. Furthermore, using the notation PT_M^S for denoting the physical computation time needed for computing all the intermediate virtual times including VT_M^S, the normalised cost for smell is defined as:

$$A_M^S = \frac{PT_M^S}{PT_{1M}^S}, \tag{10.1}$$

where the subscript "1M" indicates the highest mesh quality used in the smell transport simulations (1 million CVs). As can be seen from the above formula, the theoretical cost A_M^S is always between $[0, 1]$ while it is an increasing function of the mesh size, i.e. $A_M^{S_1} \leq A_M^{S_2}$ for meshes S_1, S_2 where S_1 has less CVs than S_2. In the work presented in this chapter, the theoretical cost for smell was estimated using the physical computation times obtained from the coarsest mesh size, therefore, formula 10.1 was used with $M = 1K$ for every scenario.

As shown in the psychophysics study ***E2*** (see chapter 9) the selection of the mesh size for simulating smell transport has no perceptual effect to the user, therefore, the selection of the smallest mesh size can be used for defining the theoretical costs for smell. These costs are given in table 10.1 for each of the four scenarios used in this study.

Scenario	Bathroom	Car	Kitchen	Kitti
VT^S_{1K}	646	458	1685	1199
PT^S_{1K}	229.2	123.6	718.8	481.3
PT^S_{1M}	6222.5	4164.7	17900.3	15228.9
A^S_{1K}	0.037	0.029	0.040	0.032

Table 10.1: All the quantities needed for computing smell theoretical costs across the four scenarios selected. The first row shows the virtual time (sec) required to reach concentration C_1. The second row shows physical computation times (sec) for the coarsest mesh (1K). The third row shows physical computation times (sec) for the finest mesh (1M). The last row gives theoretical costs for smell as ratios of the second and third row of the table.

The concentration level C_1 was decided to be delivered when the corresponding GUI control is enabled. This concentration is the first olfactory stimulus that elicits a perceptual difference starting from concentration $C_0 = 1.2$ ppm (see section 9.4) and was preferred as it is a medium intensity olfactory stimulus without being too faint or strong when sniffed. The selection of C_1 as the delivery smell impulse was based on a preliminary study implemented with six people before the main experiment. The participants of this pilot study were asked to evaluate the intensity of an olfactory stimulus using one of the adjectives "faint", "normal" or "strong". The stimuli with concentrations $C_0 = 1.2$ ppm, $C_1 = 3.4996$ ppm and $C_2 = 10.132$ ppm that were used in the psychophysics study ***E2*** were delivered to the participants using the olfactory display described in section 8.4. All the participants answered that C_1 can be characterised as a "normal" stimulus in terms of intensity.

As can be seen from table 10.1, the normalised cost for olfaction is dependent on the scenario type. The computational cost for simulating smell in a VE strongly depends on the virtual time passed until the concentration reaches C_1 ppm in the probe location. For instance, concentration C_1 is achieved much faster in the car scenario where the air draughts transfer the smell faster rather than in the Kitchen scenario where the smell plume is dispersed slowly in the room. The dependency of the theoretical cost for smell on the environmental conditions of the VE is not a phenomenon that happens inherently in the computation of the theoretical costs for visual and auditory stimuli. This dependency is a key distinction between the senses of olfaction and sight/hearing.

The same budget sizes used in the experiment ***E1*** were also used in this

study for allocating resources to the three senses. Table 7.1 gives these budgets and their notation letters.

10.3 Experimental layout

This section describes the experimental layout followed in this study. Details about the design, software and hardware materials, participant information and the experimental procedure are given.

10.3.1 Design

The experimental designed required participants to allocate a given budget by adjusting the quality of the displayed visual and auditory stimuli in a range of different scenarios. In addition, participants had the opportunity to experience an olfactory stimulus by turning ON or OFF a GUI control. Participant's preferences were captured for different budget sizes and scenarios. This design includes the effect of two independent variables, namely, *scenario* and *budget*, on three dependent variables. From the three dependent variables two are numeric denoting the percentage of the total budget that is given to adjust graphics and auditory quality and one is a categorical variable that denotes whether a participant selected (or not) to receive a smell stimulus.

The independent variable *budget* has five different levels. These are the budget sizes B_1 to B_5 that were defined in section 7.2 of the experimental study ***E1***. The independent variable *scenario* has three different levels. These are the scenarios {Bathroom,Car,Kitchen} which are a subset of the scenarios used in the experiment ***E1*** (snapshots are given in figure 7.1). One more scenario was used for training (Kitti) before the main experimental session and no data were collected from this scenario. Both of the independent variables are within participants as every subject had to repeat the task of allocating resources in every combination of budget and scenario.

10.3.2 Materials

For this study, the two monitors and the set of headphones used in the experimental study ***E1*** were also used for conducting this experiment. In addition, an olfactory display was used for delivering smell impulses when a specific GUI control was enabled by the users. The specifications of the olfactory display are described in section 8.4. A graphical representation of the hardware experimental set-up is

given in figure 10.2. The experiment was conducted in a silent room of constant ventilation.

The four scenarios were also used for conducting smell transport simulations in chapter 8 and selected because they are more congruent to the smell of citral, a VOC reminiscent of the smell of lemon, compared to the other two VEs (Yard, Restaurant). Figure 10.1 depicts the experimental scenarios and zooms in the source of smell used for the simulation of odour transport. From the user's perspective, at least one object could have been the source of smell and no information was given to the participants regarding the source of the odour.

Figure 10.1: Smell source locations of the scenarios used in the ***E3*** experimental study. From **Left** to **Right** and **Top** to **Bottom**: Bathroom, Car, Kitchen, Kitti. The circular areas show the smell source.

When the control for smell was turned on a smell impulse of 1 second was delivered to the user directly to the nose location (see figure 10.2). The delivery of smell bursts instead of a continuous flow was preferred so as to avoid user's fast adaptation to the released odour. A control channel was also used to release air bursts of 1000 ml/min flow rate and convect the smell bursts away from participant's location. Five air burst of frequency 400 ms were released when the user was pressing the "Next" or the "Smell OFF" buttons to go to the next trial or choose not to allocate resources when the "Smell ON" was enabled.

The delivery of the audio and visual stimuli was delayed by 240 ms when the smell was set on in the experiment to compensate the olfactory display's onset

time (see section 8.4). The objective was to deliver all the three sensory stimuli synchronously. The participants were not able to perceive any temporal delay of the delivered stimuli according to their answers in the feedback following the experimental session. As far as the auditory stimuli are concerned, these are the same as the audio tracks used in the experiment **E1** for the four selected scenarios.

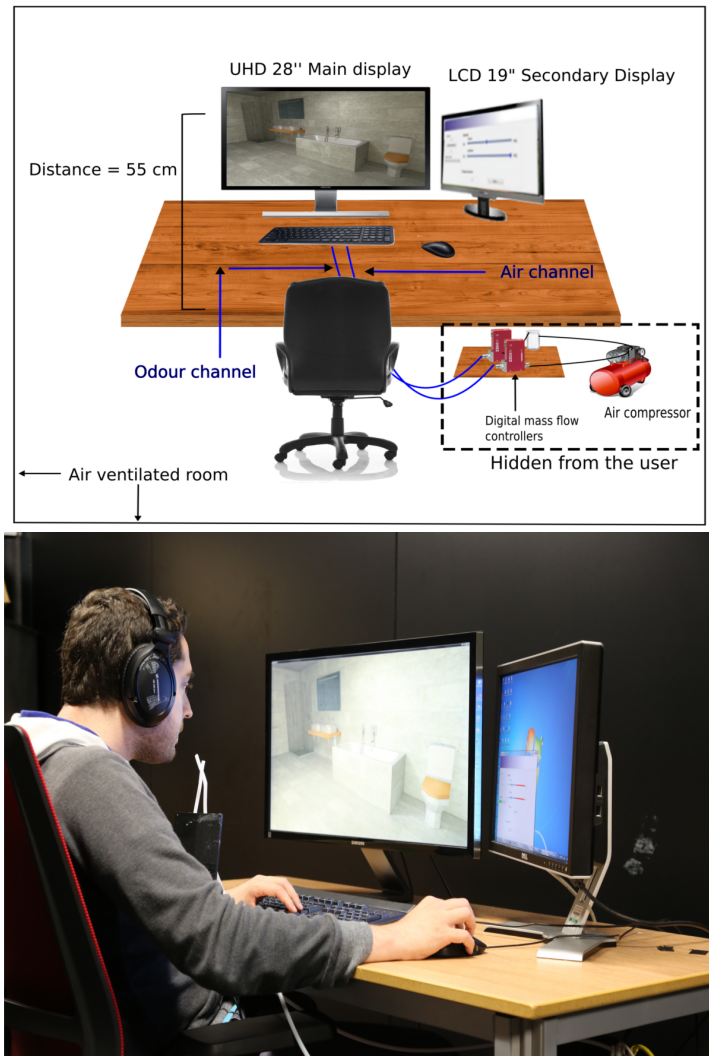


Figure 10.2: **Top:** Hardware set-up for the ***E3*** experimental study. The hardware materials for delivering visual and auditory cues are the same as the experiment ***E1***. **Bottom:** Photograph taken during preliminary tests.

A GUI application was used for carrying out the experimental study ***E3***. GUI snapshots are provided in figure 10.3. This GUI included controls for turning the smell ON or OFF depending on user's preferences. The user interface was also connected with the olfactory display using a custom API so as smell impulses to be

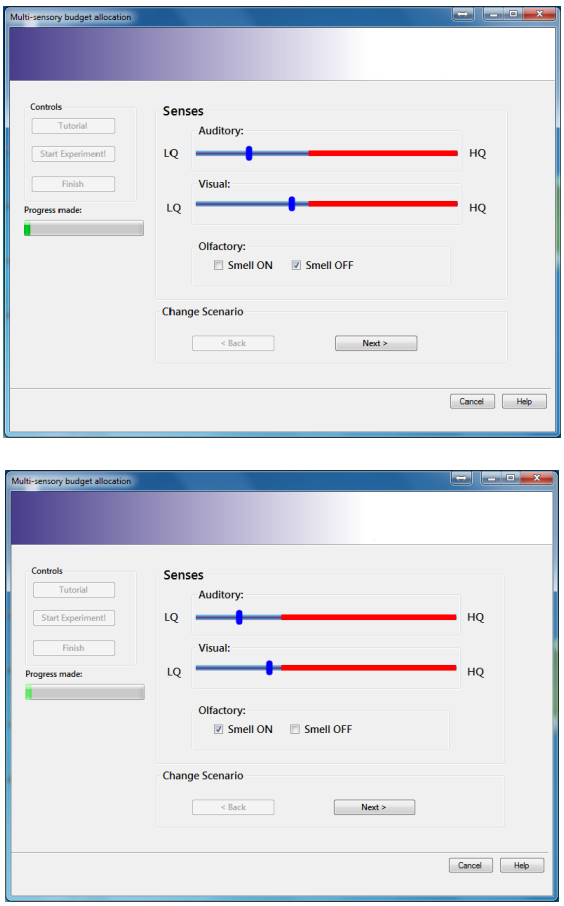


Figure 10.3: Snapshots of the software used in ***E3*** experimental study. **Top:** An instance of the budget size B_2 when the "Smell OFF" control is activated. **Bottom:** An instance of the budget size B_2 when the "Smell ON" control is activated.

released when the respective GUI control was activated by the users.

As it can be seen in figure 10.3, when the control for smell is ON, the user becomes aware of the constraint in the budget as the red coloured stripes increase in length and the number of available levels for audio-visual quality improvements are reduced. The theoretical cost for the smell stimulus is subtracted from the budget and the rest of the budget amount is used for selecting audio-visual quality.

At the beginning of each experimental trial the two sliders are positioned at the start of the slider bars which correspond to a "null" stimulus. The "null" stimulus configuration includes the delivery of a grey image and a silent track while the smell control is pre-selected OFF. The initial selection of the grey image aims to neutralise participants' eyes and is suggested by [ITU12]. When budget B_5 is available for resource allocation, the audio and visual thumbs start from a medium

quality level that has cost equal to $B_5 - B_4$. As was the case with experiment ***E1***, this initial configuration of the thumbs allows the subjects to explore all the available quality levels before deciding which quality is desired. The alternative option to start the thumbs at random positions was not preferred to avoid biasing the participant with a thumb configuration which does not represent his/her actual preferences.

At the beginning of each trial, the two thumbs are independent of each other until the sum of the theoretical costs for the three sensory stimuli exceeds the budget given for the trial. After the first attempt to exceed the budget, the thumb controls become dependent. At any point, if the user decides that the smell stimulus is required by pressing the "Smell ON" control, the GUI immediately adjusts to the new conditions and the cost for smell is subtracted from the total budget.

The transition from independent to dependent GUI controls for adjusting audio-visual quality was also applied in the experimental study ***E1*** and is the core idea of the resource allocation task. When a quality improvement is performed to either the senses of vision or hearing the quality of the other stimulus needs to be reduced accordingly so as the budget to remain always constant and equal to the current budget size. The addition of the olfactory impulses does not disrupt this mechanism as the cost for a smell impulse is immediately allocated after the control is turned ON. At this point, a reduced amount is available for distribution to the other two senses.

10.3.3 Participants

A total of 25 participants, 13 men and 12 women with ages between 23 and 46 years participated in the experiment. The average age of all the participants was 36.2 years old. Participants were coming from different academic backgrounds and had normal or corrected to normal vision. None of them reported a hearing problem as well as any temporary or permanent smell problem. The sample of participants is opportunistic as it was comprised of university students and academic staff. All the subjects participated voluntarily in the experiment and received no payment. It is important to note that subjects other than the ones participated in experiment ***E1*** were recruited for the implementation of the experimental study ***E3*** to avoid introducing bias as the two experiments use the same visual and auditory cues.

10.3.4 Procedure

Every experimental session was initiated with a tutorial where the participants had the chance to familiarise with the task of allocating resources. The Kitti scenario was used across all the budget sizes for training. During the tutorial, the experimenter had the chance to explain the objectives and purpose of the study to the participants. There was no constraint on the training time and participants could repeat any experimental trial as many times as necessary. The main experimental procedure was approximately 15 minutes. After the training session, the subjects were asked if they completely understood the task of resource allocation to the three senses and then the main experimental session was started. The subjects were asked to judge and form the best multi-sensory experience using the controls of the GUI. Appendix D includes ethical considerations for this and the other studies of this thesis.

10.4 Results

In this section the results of the experimental study ***E3*** are given. The analysis is divided into two parts for better clarity. The first part includes a repeated measures ANOVA via a 3 (*scenario*) $\times$ 5 (*budget*) factorial design for studying the effect of the independent variables on participants' decision to enable the smell impulse ON and the second part goes through a Multivariate Analysis Of Variance (MANOVA) for examining the effect of the two independent variables on the visual and auditory allocation preferences.

10.4.1 Analysis on the smell preferences

Figure 10.4 depicts the proportion of times that people preferred to receive a smell impulse along with an audio/visual stimulus for each of the three scenarios and across every budget size. As the results show, at small budget sizes, participants choose not to allocate resources for olfactory stimuli and prefer to distribute the given resources to the other two senses. As the budget increases, the frequency people select to receive an olfactory stimulus increases significantly.

Furthermore, people enable the control for smell more often in the Car and Bathroom VEs compared to the Kitchen VE. Participants' preference to these scenarios might be explained due to the ease to locate the smell source in the scene. Also, at the feedback given after the end of the experimental procedure people confirmed that were able to easily detect the smell source in the Car and Bathroom

scenes while for the Kitchen VE their answers show ambiguity of whether the smell originates from the mug or a plate with fruits next to it. The purpose of asking participants' feedback was for evaluating the experimental procedure and not for collecting extra experimental data.

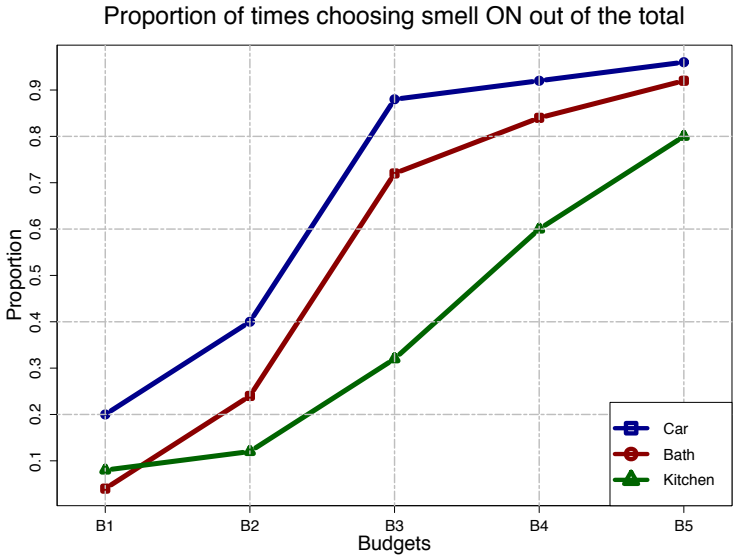


Figure 10.4: Proportion of times the subjects selected to receive a smell stimulus in each of the three scenarios and across all different levels of the factor *budget*.

The main effect of *budget* was significant with $F(4, 360) = 61.474$, $p < 0.05$ indicating a difference in the proportions the subjects selected to receive a smell impulse across the five budgets. The main effect of *budget* did not violate the assumption of sphericity (via Maulchy's test, $p > 0.05$). The main effect of *scenario* was also significant $F(2, 360) = 18.423$, $p < 0.05$ indicating a difference in the proportions the smell burst was received by the participants across the three scenarios. The main effect for *scenario* did not violate the assumption of sphericity (via Maulchy's test, $p > 0.05$). Furthermore, the interaction of *budget* $\times$ *scenario* was also examined and it was not found statistically significant, $F(8, 360) = 1.756$, $p > 0.05$. The assumption of sphericity was also not violated in this test (Maulchy's test with $p > 0.05$).

Contrast comparisons for proportions were conducted between groups of budgets using post-hoc proportion tests and applying Bonferroni correction for preserving the statistical power of the tests. These results are collected in table 10.2. The results show that small budget sizes (B_1, B_2) are not significantly different indicating that the resources are still not sufficient for distributing resources to smell. On the

other hand, large budget sizes are also not significantly different (B_4, B_5) meaning that people tend to enable the smell stimulus approximately the same number of times when distributing these budget sizes.

Scenario	**Budget size**	**p−value**
Car	B_1 B_2 B_3 B_4 B_5	< 0.05
Bath	B_1 B_2 B_3 B_4 B_5	< 0.05
Kitchen	B_1 B_2 B_3 B_4 B_5	< 0.05
All	B_1 B_2 B_3 B_4 B_5	< 0.05

Table 10.2: Contrast comparisons for proportions of smell ON between budgets at every scene and across all the scenes. Proportions of budgets with no significant differences are grouped together. The significance level of the hypothesis tests was $\alpha = 0.05$.

Contrast comparisons for proportions were also conducted between groups of scenarios using post-hoc proportion tests and applying Bonferroni correction. These results are collected in table 10.3. At small budget sizes, there is no significant difference between any of the three delivered scenarios. At larger budget sizes, the group {Car, Bath} is significantly different than the group {Kitchen} indicating that participants clearly prefer to enable the smell ON more frequently when encounter scenarios of the former group.

Budget	**Scenarios**	**p−value**
B_1	Car Bath Kitchen	< 0.05
B_2	Car Bath Kitchen	< 0.05
B_3	Car Bath Kitchen	< 0.05
B_4	Car Bath Kitchen	< 0.05
B_5	Car Bath Kitchen	< 0.05
All	Car Bath Kitchen	< 0.05

Table 10.3: Contrast comparisons for proportions of smell ON between scenarios at every budget and across all the budgets. Proportions of scenarios with no significant differences are grouped together. The significance level of the hypothesis tests was $\alpha = 0.05$.

10.4.2 Analysis on the audio-visual preferences

Figure 10.5 depicts the mean percentages and confidence intervals for graphics and acoustics quality for each scenario and across all the budgets. As these variables

are numeric, measures of central tendency and variation are meaningful. This is not the case with the smell preferences as it is a categorical variable. As can be seen from these images, the visual and auditory percentages are negatively correlated. For small budget sizes people set the visual quality as their first priority, therefore, they allocate the majority of their budget. As the budget size increases a balanced distribution of resources is preferred as the available amount permits it. It is important to note that these results confirm the outcome of the earlier experimental study ***E1*** (see section 7.4). The effect of different scenarios in resource allocation is more evident in graphics as the curves are more distant to each other compared to the curves of the plot in acoustics.

MANOVA analysis was applied to examine the effect of the *budget* and *scenario* on the audio-visual allocation percentages. Specifically, the main effect of *budget* was significant for the percentage devoted to audio quality with $F(4, 360) = 153.85$, $p < 0.05$ while the effect of *scenario* was not found statistically significant with $F(2, 360) = 0.0241$, $p > 0.05$. The main effect of *budget* and *scenario* in these tests did not violate the assumption of sphericity (via Maulchy's test, $p > 0.05$). It should be noted that the factor *scenario* was found significant in the experimental study ***E1*** where the scenarios included both human and mechanical sound stimulus. In this study, only scenarios with mechanical sound were used. The interaction *scenario* $\times$ *budget* was not found significant, $F(8, 360) = 0.012$ $p > 0.05$, with no violation of the assumption of sphericity (via Maulchy's test, $p > 0.05$). The main effect of *budget* was significant for the percentage devoted to graphics with $F(4, 360) = 42.116$, $p < 0.05$. The effect of *scenario* was also found statistically significant for the graphics allocation percentage with $F(2, 360) = 11.032$, $p < 0.05$. The main effect of both factors did not violate the assumption of sphericity (via Maulchy's test, $p > 0.05$). The interaction *scenario* $\times$ *budget* was not found significant, $F(8, 360) = 0.9376$, $p > 0.05$. The assumption of sphericity was also not violated (via Maulchy's test, $p > 0.05$).

Contrast comparisons using post-hoc t-tests were also conducted for investigating groups of budgets that are not significantly different across the scenarios and also for finding groups of scenarios that are not significantly different across budgets. These contrasts are given in tables 10.4 and 10.5 for graphics and acoustics respectively.

As can be seen from the last rows of the tables in 10.4, small budget sizes (B_1, B_2) are not significantly different meaning that participants follow similar allocation strategy for graphics when given these budgets independently of the scenario. The same argument also holds for very large budget sizes (B_4, B_5). As far as the

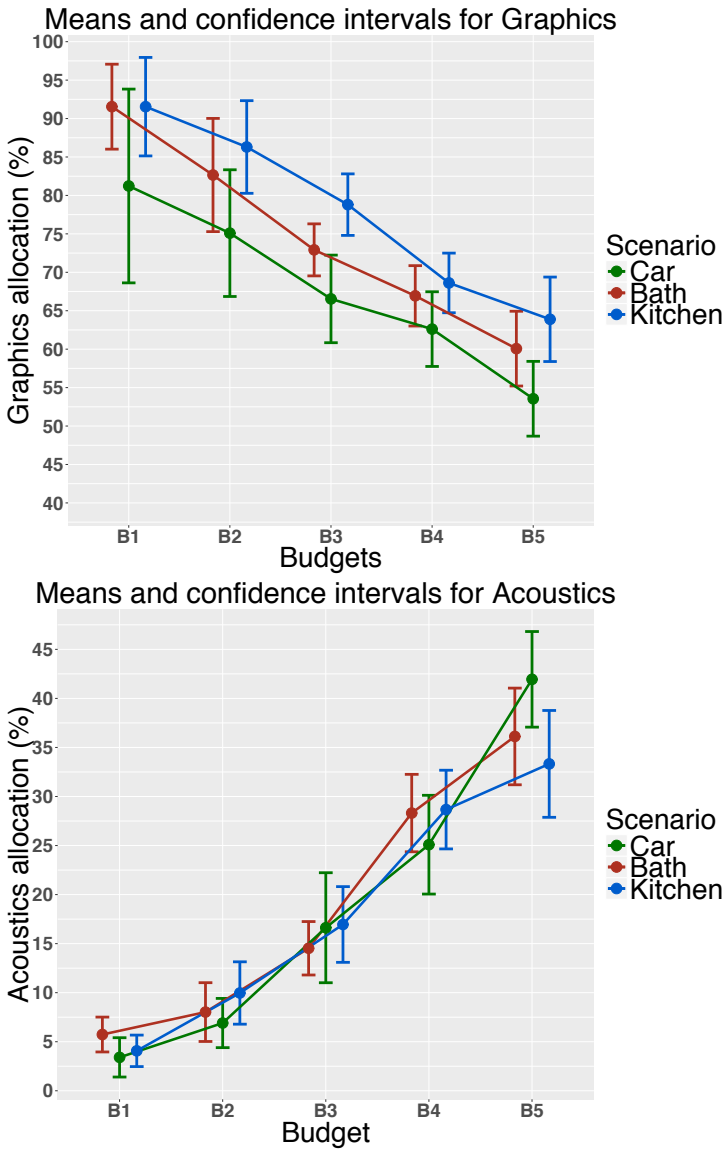


Figure 10.5: Mean allocation percentages and confidence intervals for every scenario and across all the budget sizes. Jittering has been applied to the two plots for better visualisation of the results. **Top:** Graphics, **Bottom:** Acoustics.

scenarios are concerned, subjects tend to increase graphics quality in a similar way in the scenarios Kitchen and Bath.

Table 10.5 shows that none of the scenarios is significantly different from the others when distributing resources for aural quality improvements (table on right hand side). As far as the budgets are concerned, the same trend for graphics holds also for the acoustics for the effect of the budget size. People do not change their allocation approach at very small or very large budget sizes (table on left hand side).

Scenario	Budget size					p−value
Car	B_1	B_2	B_3	B_4	B_5	< 0.05
Bath	B_1	B_2	B_3	B_4	B_5	< 0.05
Kitchen	B_1	B_2	B_3	B_4	B_5	< 0.05
All	B_1	B_2	B_3	B_4	B_5	< 0.05

Budget	Scenarios			p−value
B_1	Car	Bath	Kitchen	< 0.05
B_2	Car	Bath	Kitchen	< 0.05
B_3	Car	Bath	Kitchen	< 0.05
B_4	Car	Bath	Kitchen	< 0.05
B_5	Car	Bath	Kitchen	< 0.05
All	Car	Bath	Kitchen	< 0.05

Table 10.4: **Left:** Contrast comparisons between budgets at every scenario and across all the scenarios for graphics, **Right:** Contrast comparisons between scenarios at every budget and across all the budgets for graphics. Budgets or scenarios with no significant difference are grouped together. The significance level of the hypothesis tests was $\alpha = 0.05$.

Scenario	Budget size					p−value
Car	B_1	B_2	B_3	B_4	B_5	< 0.05
Bath	B_1	B_2	B_3	B_4	B_5	< 0.05
Kitchen	B_1	B_2	B_3	B_4	B_5	< 0.05
All	B_1	B_2	B_3	B_4	B_5	< 0.05

Budget	Scenarios			p−value
B_1	Car	Kitchen	Bath	< 0.05
B_2	Car	Kitchen	Bath	< 0.05
B_3	Car	Kitchen	Bath	< 0.05
B_4	Car	Kitchen	Bath	< 0.05
B_5	Car	Kitchen	Bath	< 0.05
All	Car	Kitchen	Bath	< 0.05

Table 10.5: **Left:** Contrast comparisons between budgets at every scenario and across all the scenarios for acoustics, **Right:** Contrast comparisons between scenarios at every budget and across all the budgets for acoustics. Budgets or scenarios with no significant difference are grouped together. The significance level of the hypothesis tests was $\alpha = 0.05$.

10.5 Estimation models

All the experimental data were used in the construction of two statistical models, denoted as $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ throughout this thesis. The former statistical model takes only the budget as input and it gives estimations independent of the given scenario while the latter takes as inputs both the budget and the scenario i.e. one of the {Bathroom, Car, Kitchen}. Both of these models are assessed later with actual human allocation preferences gathered from a validation experimental study. For the validation, untested budgets and an untested scenario were used to gather humans' allocation preferences. This section begins with the description of the two models and later demonstrates how the two models perform against the data collected from the validation stage.

10.5.1 Models

The proposed statistical models are used to estimate three quantities. The first quantity is a probability that expresses whether smell should be given for an input budget and scenario while the other two estimations are proportions of the total budget that should be devoted for improving graphics and acoustics quality. In the following, the description of the two models is given in separate subsections.

Model $\mathbf{M_1^{E3}}$:

As pointed out in the previous section, the decision to turn the smell stimulus ON or OFF is a categorical variable thus, a logistic regression was used to model it. Specifically, if the probability to give a smell impulse is denoted as p then the ratio $\frac{p}{1-p}$ expresses the odds of enabling the release of the smell stimulus. Then the logistic regression model for smell can be written as:

$$\log\left(\frac{p}{1-p}\right) = \hat{\beta}_i^s + \hat{\beta}_b^s \cdot \text{budget}, \tag{10.2}$$

where the coefficients $\hat{\beta}_i^s$ and $\hat{\beta}_b^s$ are the least squares regression estimates. The subscripts "i" and "b" are used as shorthands for the intercept and the budget respectively. When a budget size is given as input, the right hand side of the formula 10.2 is evaluated and if the result is positive then it can be inferred that $p > 1-p$ and thus a smell impulse should be delivered to the user. The above model is used only to estimate whether the smell should be turned ON/OFF and is one component of the whole model $\mathbf{M_1^{E3}}$. The full model is also composed of two more components for estimating audio-visual allocation percentages. These components are estimated using linear regression models. A formula for the full multivariate model $\mathbf{M_1^{E3}}$ can be written in matrix form as follows:

$$\begin{bmatrix} \hat{S}_1 \\ \hat{V}_1 \\ \hat{A}_1 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_i^s & \hat{\beta}_b^s \\ \hat{\beta}_i^v & \hat{\beta}_b^v \\ \hat{\beta}_i^a & \hat{\beta}_b^a \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \text{budget} \end{bmatrix}, \tag{10.3}$$

where the letters $\hat{V}_1$ and $\hat{A}_1$ are the vision and audio allocation estimations while $\hat{S}_1 = \log\left(\frac{p}{1-p}\right)$ is the logistic regression component of the model for smell as given in formula 10.2. Hypothesis tests of the form $H_0 : \beta_z^k = 0$ vs $H_a : \beta_z^k \neq 0, \quad z \in \{i, b\}$ and $k \in \{s, v, a\}$ were conducted to check the significance of the unknown paameters $\beta_i^s, \beta_b^s, \beta_i^v, \beta_b^v, \beta_i^a, \beta_b^a$. In all these tests the null hypothesis is rejected indicating that the respective least squares regression estimators $(\hat{\beta}_i^s, \hat{\beta}_b^s, \hat{\beta}_i^v, \hat{\beta}_b^v, \hat{\beta}_i^a, \hat{\beta}_b^a)$

should be remained in the model formulation. The six estimates on the right hand side of the above equation are given in table 10.6.

The coefficients of determination R^2 and R^2_{adj} for the goodness of fit for model $\mathbf{M_1^{E3}}$ are given in table 10.7. The superscript in parenthesis is used to denote the sense. As the smell estimation is based on logistic regression the first coefficient of determination is based on the Akaike Information Criterion (AIC) (the lower the better) while the other two for graphics and audio are the Nagelkerke coefficients of determination (the higher the better).

Model	$\hat{\beta}_i^s$	$\hat{\beta}_b^s$	$\hat{\beta}_i^v$	$\hat{\beta}_b^v$	$\hat{\beta}_i^a$	$\hat{\beta}_b^a$
$\mathbf{M_1^{E3}}$	-1.31	0.03	84.89	-0.25	4.72	0.32

Table 10.6: Least squares regression estimates of the multivariate model $\mathbf{M_1^{E1}}$. The subscripts "i" and "b" are used as shorthands for intercept and budget respectively. The superscipts "s", "v", "a", are used instead of smell, vision and audio respectively.

Model	$R^{2(s)}$	$R^{2(v)}$	$R^{2(v)}_{\text{adj}}$	$R^{2(a)}$	$R^{2(a)}_{\text{adj}}$
$\mathbf{M_1^{E3}}$	24.8	0.3647	0.3627	0.61	0.609

Table 10.7: Coefficients of determination for the multivariate model $\mathbf{M_1^{E3}}$. The akaike information criterion was used for the $R^{2(s)}$. The Nagelkerke coefficient is used for the coefficients of determination for vision and audio.

Model $\mathbf{M_2^{E3}}$:

Model $\mathbf{M_2^{E3}}$ takes into account not only the budget size but also the scenario. For different scenario inputs the model adapts its estimate accordingly. The logistic regression component of the model can be written as follows:

$$\log\left(\frac{p}{1-p}\right) = \hat{\gamma_i^s} + \hat{\gamma_b^s} \cdot \text{budget} + \mathbb{1}_{\text{C}} \cdot \hat{\gamma_{\text{C}}^s} + \mathbb{1}_{\text{K}} \cdot \hat{\gamma_{\text{K}}^s}, \tag{10.4}$$

where the term $\mathbb{1}_{\text{C}} \cdot \hat{\gamma_{\text{C}}^s}$ is added when the scenario Car is given as input (subscript "C") while the term $\mathbb{1}_{\text{K}} \cdot \hat{\gamma_{\text{K}}^s}$ is added when the scenario Kitchen is given as input (subscript "K"). When the Bathroom scenario is given, both of the indicator functions $\mathbb{1}_{\text{C}}$ and $\mathbb{1}_{\text{K}}$ result to 0 and the estimate is based only on the coefficients $\hat{\gamma_i^s}$ and $\hat{\gamma_b^s}$. The visual-audio allocation percentages are estimated using linear regression estimates as previously. The full multivariate regression model $\mathbf{M_2^{E3}}$ can be written

in matrix form as follows:

$$\begin{bmatrix} \hat{S}_2 \\ \hat{V}_2 \\ \hat{A}_2 \end{bmatrix} = \begin{bmatrix} \hat{\gamma_i^s} & \hat{\gamma_b^s} & \hat{\gamma_{\mathrm{C}}^s} & \hat{\gamma_{\mathrm{K}}^s} \\ \hat{\gamma_i^v} & \hat{\gamma_b^v} & \hat{\gamma_{\mathrm{C}}^v} & \hat{\gamma_{\mathrm{K}}^v} \\ \hat{\gamma_i^a} & \hat{\gamma_b^a} & \hat{\gamma_{\mathrm{C}}^a} & \hat{\gamma_{\mathrm{K}}^a} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \text{budget} \\ \mathbb{1}_{\mathrm{C}} \\ \mathbb{1}_{\mathrm{K}} \end{bmatrix}, \tag{10.5}$$

where the letters $\hat{V}_2$ and $\hat{A}_2$ are used to denote the vision and audio estimations of this model while $\hat{S}_2$ is the logistic regression component as given in formula 10.4.

It can be seen from the above that model $\mathbf{M_1^{E3}}$ is nested to the model $\mathbf{M_2^{E3}}$ and it is important to see whether the six extra unknown parameters γ_C^s, γ_C^v, γ_C^a, γ_K^s, γ_K^v, γ_K^a of the estimators $\hat{\gamma_C^s}, \hat{\gamma_C^v}, \hat{\gamma_C^a}, \hat{\gamma_K^s}, \hat{\gamma_K^v}, \hat{\gamma_K^a}$ are statistically significant and should be remained in the model $\mathbf{M_2^{E3}}$. Parameters that are not significant indicate that their estimators have no contribution to the model predictions and should be removed. Hypothesis tests of the following form:

$$H_0 : \gamma_i^k = 0 \text{ vs } H_a : \gamma_i^k \neq 0, \quad i \in \{C, K\} \text{ and } k \in \{s, v, a\}, \tag{10.6}$$

were conducted to see the significance of these parameters. It was found that for the parameter γ_K^v the H_0 is not rejected ($p > 0.05$) while the same also holds for the unknown parameters γ_K^a, γ_C^a. The latter parameters were expected to be not significant (not reject the H_0 hypothesis) as the factor *scenario* was not found to be statistically significant for adjusting aural quality in the MANOVA tests of the previous section. The least square regression estimates of the $\mathbf{M_2^{E3}}$ model are given in table 10.8.

Model: $\mathbf{M_2^{E3}}$	Estimates			
Sense	intercept	budget	Car	Kitchen
Smell	-1.35	0.04	0.72	-1.10
Visual	87.33	-0.25	-7.29	$*$
Audio	4.72	0.32	$*$	$*$

Table 10.8: Least squares regression estimates of the multivariate model $\mathbf{M_2^{E1}}$ given in formula 10.5. The asterisks show the parameters that are not statistically significant (not reject the H_0 hypothesis) and their estimators were excluded from the statistical model.

The coefficients of determination R^2 and R^2_{adj} for the goodness of fit for the final form of the model $\mathbf{M_2^{E3}}$ are given in table 10.9. The superscript in parenthesis is used to denote the sense. As the smell estimation is based on logistic regression

the first coefficient of determination is based on the akaike information criterion (AIC) (the lower the better) while the other two for graphics and audio are the Nagelkerke coefficients of determination (the higher the better).

Model	$R^{2(s)}$	$R^{2(v)}$	$R^{2(v)}_{\text{adj}}$	$R^{2(a)}$	$R^{2(a)}_{\text{adj}}$
$\mathbf{M_2^{E3}}$	18.4	0.40202	0.3985	0.654	0.62

Table 10.9: Coefficients of determination for the multivariate model $\mathbf{M_2^{E3}}$. The akaike information criterion was used for the $R^{2(s)}$. The Nagelkerke coefficient is used for the coefficients of determination for vision and audio.

10.6 Validation

The performance of the two proposed statistical models was compared with data collected from a validation experimental study. This study included the use of untested budget sizes and a new untested scenario. The objective was to test the performance of the two models with inputs other than the ones used for their construction. The validation study follows the same guidelines as the main experimental study ***E3***. In the following the experimental method and the results are discussed.

10.6.1 Design

The experimental layout of the validation study follows from the experimental procedure described in section 10.3. For this experiment three scenarios were used, namely, Car, Kitchen and Kitti. The scenario Kitti was used in the main experiment for training, therefore none of the models was constructed from data collected from this VE. Three budget sizes were used in this experiment, each one different from the previously used budgets. The three budgets were selected in the midpoints between B_1 and B_2, B_2 and B_3 and B_3 and B_4. These are denoted as NB_1, NB_2 and NB_3 in the following. All the materials of the main study, were also used for the validation stage with no change. As the number of budgets is limited to three, each participant had to undertake nine experimental trials in the main session and three trials for training. The Kitchen scenario was used for training.

A total of six people, three men and three women, participated and the average age of all the subjects was 32.2 years old. None of these participants was familiar with the objectives of the experiment as they had not taken part in any of the previous studies ***E1*** or ***E3***. All the participants had normal or corrected to normal vision and reported no hearing or smelling problems.

10.6.2 Results

The two models were compared with the data collected from the validation study to investigate their performance against real human preferences. As mentioned in section 10.5 the parameters γ_K^a, γ_C^a were not found statistically significant (not reject the H_0 hypothesis) therefore their estimators are not included in the model $\mathbf{M_2^{E3}}$. This was expected as the factor *scenario* was not significant for the percentage devoted to acoustics, therefore the two models give the same estimations for audio allocation.

Figure 10.6 shows the means and confidence intervals for the car scenario. The model estimations are also depicted as coloured curves for comparison. For graphics, the average error for absolute difference for the model $\mathbf{M_1^{E3}}$ is 5.8% while for the model $\mathbf{M_2^{E3}}$ is 2.3% for the car scenario. For acoustics, both models give an average error of absolute difference of 3.4%. For graphics, the model $\mathbf{M_2^{E3}}$ adapts for the car scenario, therefore only the indicator $\mathbb{1}_{\mathrm{C}}$ takes the unity value. For small budgets (NB_1) both models predict equally well for graphics quality while as the budget size increases model $\mathbf{M_2^{E3}}$ gives more accurate estimations.

Figure 10.7 shows the proportion of times smell was selected ON by the participants in the car scenario for each of the three test budgets. In the same figure, the probabilities to turn the smell ON (sigmoid curves) as predicted by the logistic regression components of the two models are depicted. As is the case with graphics, the model $\mathbf{M_2^{E3}}$ adapts its estimations for the scenario car. As can be seen, the effect of the scenario type is evident in people's preferences and model $\mathbf{M_2^{E3}}$ estimates this behaviour better.

Figure 10.8 shows the means and confidence intervals for the Kitchen scenario for the data collected during the validation stage. Model $\mathbf{M_2^{E3}}$ adjusts its estimations for this scenario by setting the indicator function $\mathbb{1}_{\mathrm{C}}$ equal to 0. For graphics, the average error for absolute difference is 4.4% and 1.6% for the models $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ respectively. For acoustics, the two models give an average error of absolute difference of 2.3% from the real data. Figure 10.11 shows the proportion of times smell was selected ON by the subjects in the Kitchen scenario. As can be seen from the three figures, at small budget sizes people prefer a low quality audio while smell is not to be ON. For the test budgets NB_1 and NB_2, the two models overestimate for audio and smell while they underestimate for graphics.

Figure 10.10 shows the means and confidence intervals for the experimental data of the Kitti scenario.This scenario was used only for training in the main experiment ***E3*** and none of the models was constructed from data coming from this scenario. People, in general, prefer high auditory quality as the budget size increases

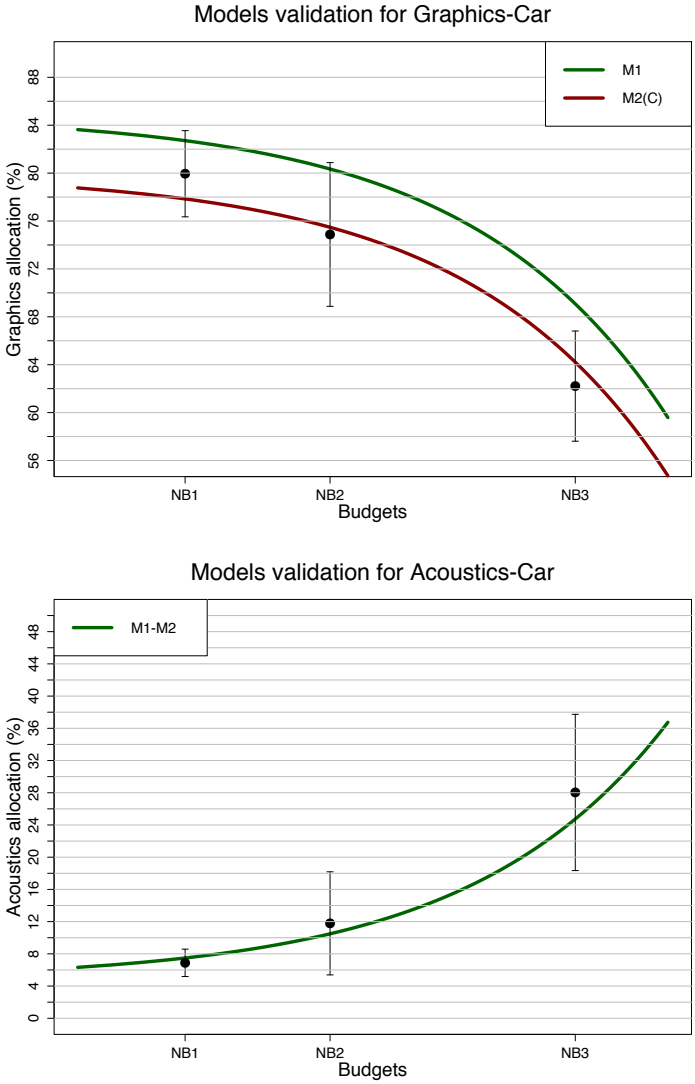


Figure 10.6: Means and confidence intervals for the Car scenario are depicted in black colour. The green and red curves are the models $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ respectively. Log-spacing was applied to the x-axis for better visualisation. **Top:** Comparison results for graphics. **Bottom:** Comparison results for acoustics.

for this scenario, a result that is confirmed also from the experimental study ***E1***. As can be seen, participants' preferences for graphics are closer to the estimations of the $\mathbf{M_2^{E3}}$ for car (i.e. $\mathbb{1}_{\mathrm{C}} = 1$). Specifically, for graphics the average error for absolute difference is 1.6% and 5.8% for the models $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ respectively while for acoustics the two models give an average error of absolute difference of 4.9% indicating that participants' need for acoustic quality was significantly increased

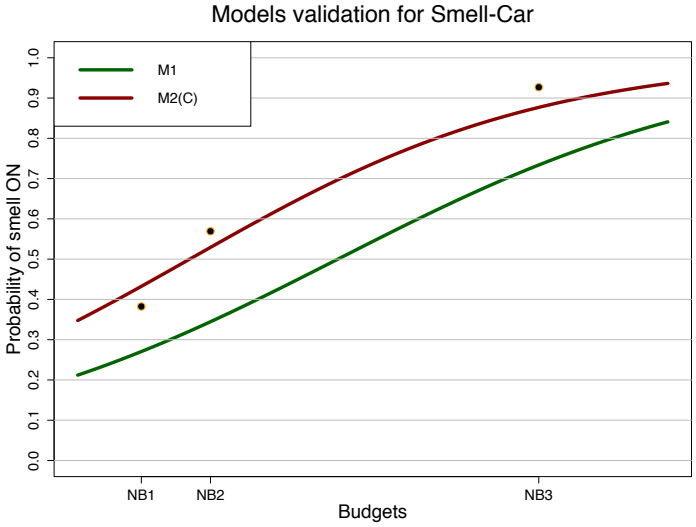


Figure 10.7: The dots represent the proportion of times smell was preferred ON by the participants for each of the test budgets in the Car scenario. The green and red curves are the probability estimations to turn the smell ON for the models $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ respectively.

compared to the estimates given by the two models.

As far as the smell for the Kitti scenario is concerned, people tended to switch the smell ON quite frequently. Model $\mathbf{M_2^{E3}}$ for the scenario Car described earlier in this section gives estimations closer to the real data compared to the model $\mathbf{M_1^{E3}}$. It should be noted that the smell source in both these scenarios {Car,Kitti} was quite easy to localise according to participants' answers in the feedback session followed the experiment. In general, participants preferred to receive the smell impulse more times in these two scenarios compared with the Kitchen scenario.

10.7 Discussion

The results of the experimental study ***E3*** yield a number of potentially useful findings related to the way humans tend to allocate resources for the multi-sensory VEs considered in this chapter. These results complement the conclusions extracted from the resource allocation study in chapter 7. First of all, the availability of more than two senses in the experimental set-up does not affect humans' general trend to devote the majority of their budget for improving the visual quality compared to the other sensory stimuli

For the scenarios used, people's priority is to obtain a clear visual stimulus

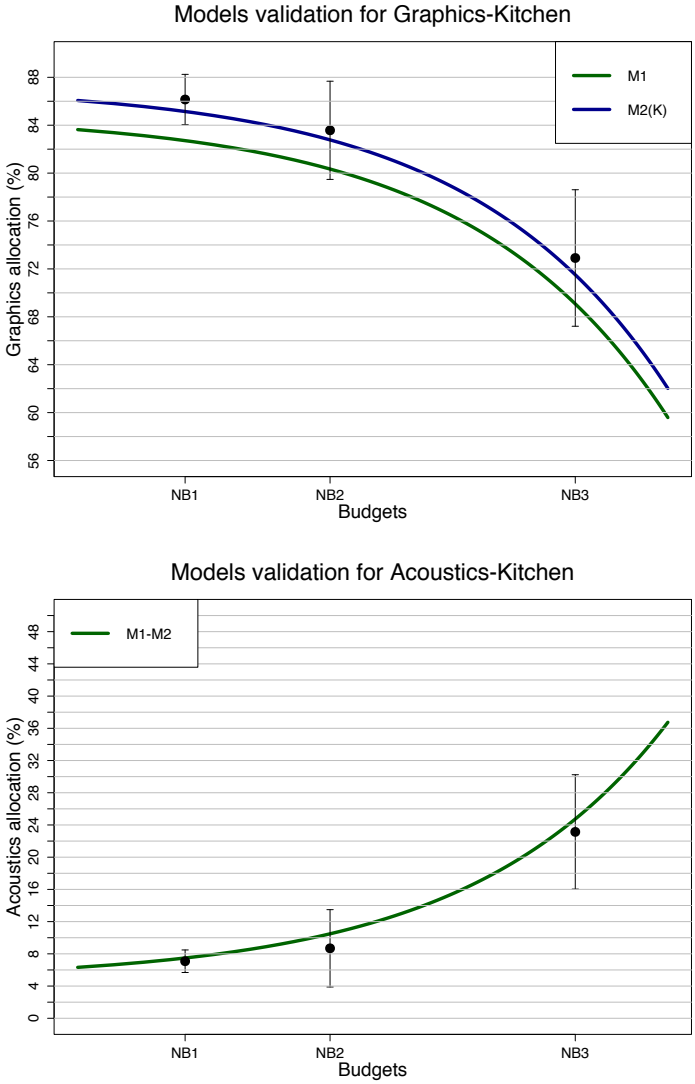


Figure 10.8: Means and confidence intervals for the Kitchen scenario are depicted in black colour. The green and blue curves are the models $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ respectively. Log-spacing was applied to the x-axis for better visualisation.**Top:** Comparison results for graphics. **Bottom:** Comparison results for acoustics.

while compromises to the audio quality and the activation of the smell stimulus are preferred when small budget sizes are available. As more resources are given, people prefer an approximately balanced distribution of resources while the frequency of times the smell stimulus is set ON indicates the importance of the sense of olfaction in the resource allocation task. Participants seemed to appreciate the addition of the olfactory impulses along with the audio-visual stimuli.

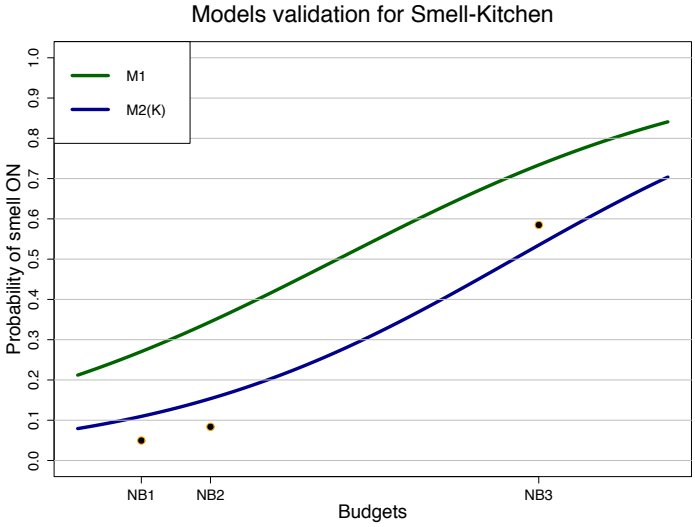


Figure 10.9: The dots represent the proportion of times smell was preferred ON by the participants for each of the test budgets in the Kitchen scenario. The green and blue curves are the probability estimations to turn the smell ON for the models $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ respectively.

As far as the effect of the scenario is concerned, it seems that the four scenarios affect people's decision to enable the smell impulse release. People might turn the smell ON more frequently in the scenarios where the source of smell was unambiguously visible in the scene. This outcome requires validation through further investigation where the importance of the smell stimuli will be judged depending on the ability to identify the source (or sources) of smell in the visual field. The scenario selection was also important for the visual allocation preferences while it has no effect in the percentage devoted for aural quality improvements.

In general, participants had no concern to favour any of the four scenarios more than others in terms of the acoustic quality. As it was shown from the results of the ***E1*** experimental study, the subset of VEs that contained human voice or melody were assigned higher acoustic quality (see chapter 7). This result can be explained due to the use of scenarios that contained no-human voice in the ***E3*** experiment. In the validation study, participants allocate higher percentage of the given budget to the auditory quality of the Kitti scenario indicating a clear intention to hear clearly the contents of the audio track.

The validation experiment shows evidence that the statistical models provide allocation estimates of different accuracy, especially in the case of the visual percentages. Model $\mathbf{M_1^{E3}}$ gives relatively accurate predictions but it was outperformed

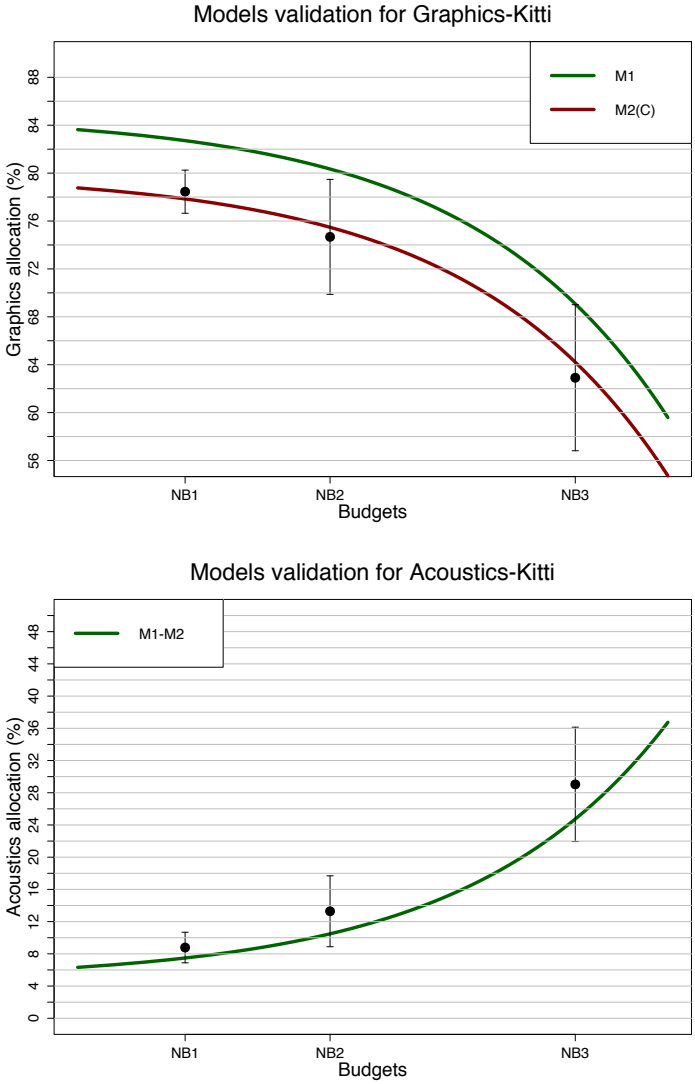


Figure 10.10: Means and confidence intervals for the Kitti scenario are depicted in black colour. The green and red curves are the models $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ respectively. Log-spacing was applied to the x-axis for better visualisation.**Top:** Comparison results for graphics. **Bottom:** Comparison results for acoustics.

by model $\mathbf{M_2^{E3}}$ in almost all the combinations of budget-scenario test conditions. The results of the validation study indicate also that the two models give relatively accurate estimations for the three different budgets and the extra scenario. The new budgets and scenario have not been used for their construction, meaning that these models could be used for testing in different experimental conditions.

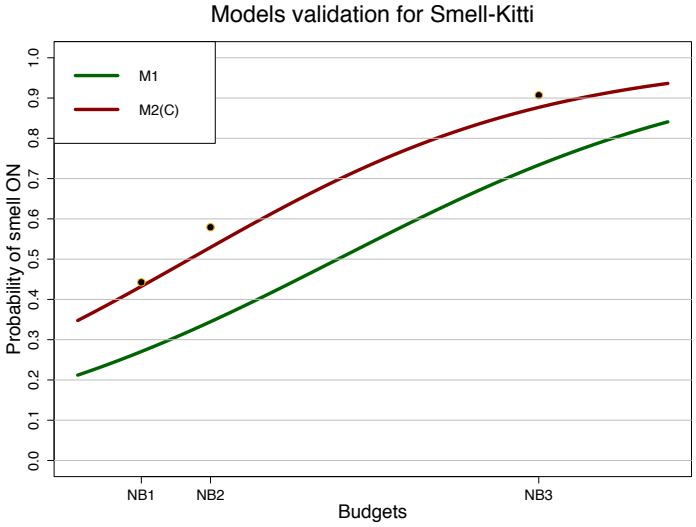


Figure 10.11: The dots represent the proportion of times smell was preferred ON by the participants for each of the test budgets in the Kitti scenario. The green and red curves are the probability estimations to turn the smell ON for the models $\mathbf{M_1^{E3}}$ and $\mathbf{M_2^{E3}}$ respectively.

10.7.1 Limitations

The work presented in this chapter certainly has limitations that can be considered in future experimental frameworks and resource allocation schemas. The delivery of the visual stimuli as static images is the main limitation of this experimental study as it restricts the number of possible applications of the proposed estimation models. Both the auditory and olfactory stimuli have an inherent temporal dimension (duration of the audio track/smell burst) that is not exploited with the use of static visual images. However, the experimental methodology presented in this chapter and chapter 7 provide useful insights on extending this work using animations or interactively changing environments.

Another possible limitation is the use of the binary (ON/OFF) scale for olfaction. It can be said that the two-level methodology for delivering smell impulses restricts user's available options but it is not clear whether (and how) a hypothetical multiple quality scale for smell would benefit user's virtual experience, particularly with the current olfactory hardware. The existence of multiple smell stimuli, on the other hand, is considered more important in a multi-sensory VE set-up and can admittedly increase the level of immersion as it is closer to what happens in reality. The work presented in this chapter is limited to the delivery of a single odorous VOC; however, the methodology for smell transport simulations and the estimation

of the JND for odour intensity (see chapters 8 and 9 respectively) can be applied to an arbitrary number of VOCs or combinations of them.

The use of the normalised costs instead of physical computational costs can also be considered a limitation. The merits of working with normalised theoretical costs have been discussed previously in sections 7.2 and 7.7.1 of this thesis.

10.8 Summary

This chapter presented an experimental study ($N = 25$) where participants had the task to allocate a given budget of resources for improving the visual and audio quality in four virtual scenarios. Participants had also the chance to enable the release of an olfactory stimulus in addition to performing audio-visual quality improvements. The sense of olfaction included two different quality levels (smell ON/OFF) while smell impulses were delivered to the user using an olfactory display. When smell is decided ON by the user, a portion of the total budget is allocated for the smell stimulus while the rest of it is available for audio-visual quality improvements. In case the smell is preferred OFF, the whole budget is available for distribution to the other senses. The results demonstrate that people clearly favour higher graphics quality in small budget sizes but tend to prefer a more balanced distribution of the resources once higher budgets become available for the scenarios tested and presented. Also, humans tend to activate the smell stimulus more frequently for larger budgets. Based on the results, a multivariate regression model is proposed. This model estimates percentages that should be given for audio-visual quality improvements and, additionally, whether smell impulses should be delivered to the user along with the other senses.

Chapter 11

Conclusions and Future Work

This chapter concludes the work of this thesis. An overview of the findings presented in the previous chapters is given and ideas are proposed for extending this work in the future. The limitations of the current work are also collected for the reader's reference.

Cross modal interactions between the senses have been successfully used in the field of Computer Graphics for alleviating the large computational times without quality degradation being perceived by users. The majority of previous studies have concentrated on interactions between sight and hearing as these are the main senses that contribute to form a basic cognition of an environment. However, humans depend on all the senses to get a full perception of the world thus sensory stimuli such as tactile, gustatory, olfactory or even the feeling of pain, temperature variation, acceleration, etc have their own importance for increasing the level of realism in virtual worlds. The introduction of stimuli other than visual and auditory has recently began to appear and it will continue increasing in future applications.

Previous work has concentrated mostly in using the effect of audio-visual cross-modal interactions as guidelines for reducing the computational complexity of one of the sensory stimuli but the level of this interaction was not really understood when available computing resources change. This thesis has presented a series of experimental studies for investigating the importance of all the sensory stimuli in a multi-sensory VE set-up where the available resources and the context of the VE vary. Specifically, in chapter 7 the problem was approached in bi-modal VEs where auditory and visual sensory stimuli were delivered to the participants. The way the experimental framework was constructed allows the introduction of any number of sensory cues as long as there is suitable infrastructure to simulate and deliver the diverse stimuli. As an example, this thesis moved towards the introduction

of olfactory impulses along with the senses of hearing and sight to the users in a tri-modal set-up (see chapter 10).

This chapter has the following structure: sections 11.1, 11.2 and 11.3 provide a brief overview of the findings presented in chapters 7, 9 and 10 respectively. Section 11.4 enumerates the limitations of this work while ideas to extend the present work are given in section 11.5. Final remarks are presented in section 11.6.

11.1 Resource allocation for bi-modal virtual environments

The objective of this reseearch was to examine how the budget of available resources and the virtual scenario affect the perceived fidelity of an audio-visual experience in a bi-modal VE. The experimental framework included the application of normalised budget sizes and costs for audio visual interactions to avoid the problem of large discrepancies in physical computation times between a digital image and a RIR. Cost normalisation makes the problem independent of the algorithm selection and its parameters while it allows to study the importance of each sense with respect to the budget size. When the quality of one sensory stimulus is increased, the quality of the other is necessarily decreased so the budget size remains constant and equal to one of the five budget sizes. The experimental data are used in the construction of two estimation models that are validated with actual human allocation preferences collected from a subsequent experiment.

For the scenarios and budget sizes used, the results showed that both the budget size ($F(2.27, 77.2) = 37.5$, $p < 0.05$) and the scenario ($F(1.87, 63.3) = 7.27$, $p < 0.05$) are statistically significant. At small budget sizes people prefer high visual quality across all the scenarios while as the budget increases a balanced distribution of resources is preferred (see figure 7.4). The effect of the scenario is quite evident in allocation preferences as the contrast comparison between the groups of scenarios {Restaurant, Kitti} and {Kitchen, Car, Yard} indicated that people clearly favour a higher aural quality in the scenarios of the former group ($F(4, 874) = 5.092$, $p = 0.015 < 0.05$) meaning that the content of the stimulus (human voice/artificial audio) affects the way people allocate resources in those scenarios. Moreover, the results of the validation study showed the prediction model that is based on both the scenario type and budget size ($\mathbf{M_2^{E1}}$) gives estimations closer to the collected experimental data. Specifically, the error of absolute difference for the latter model was 0.69%.

11.2 JND threshold for perceptually assessing smell simulations in virtual environments

The aim of the psychophysics experiment described in chapter 9 was to estimate consecutive JNDs for citral intensity in the range of concentration values provided by an available olfactory display. This experiment serves a twofold purpose. Firstly, to investigate how many different intensity levels for the smell of citral were perceptually distinguishable, on average. Secondly, to assess whether better spatial discretisation of the computational domain in smell transport simulations elicits perceptual differences that are greater than the estimated JND. Chapter 8 explained the process of simulating smell transport in four VE case studies via CFD techniques. Concentration values were captured using virtual probes located in the assumed positions of the human viewer/listener in the VE.

Two JND thresholds were estimated in the range $I = [C_0, C_{\max}] = [1.2, 11.2]$ ppm. These were $\text{JND}_1 = 2.2996$ ppm and $\text{JND}_2 = 6.6324$ ppm while the corresponding concentration stimuli that elicit the perceptual differences in the interval I are $C_1 = 3.4996$ ppm starting from $C_0 = 1.2$ ppm as pedestal stimulus and $C_2 = 10.132$ ppm starting from C_1 as pedestal stimulus.

The maximum difference between the concentration values recorded at the coarsest (1K) and the finest discretisation level (1M) during the smell transport simulations were: 1.981 ppm for the Bathroom scenario, 1.623 ppm for the car cabin scenario, 1.774 ppm for the Kitchen scenario and 1.345 ppm for the Kitti scenario (see also figure 8.10). These results indicate that excessively fine discretisation levels of the computational mesh for the scenarios presented in section 8.3 have no perceptual impact to the users, on average.

11.3 Resource allocation for tri-modal virtual environments

Chapter 10 extends the experimental methodology for resource allocation proposed in ***E1*** to include the delivery of smell impulses using an olfactory display. This experiment was based on the methodology followed in chapter 7. In addition to the sliders used for adjusting audio-visual quality, the GUI provides controls for turning the smell ON or OFF in the experimental set-up. This experiment makes use of the normalised costs for visual and audio cues while theoretical costs for the smell stimuli are also introduced based on the physical computational times recorded from the smell transport simulations in section 8.3.2. Based on the collected data of this

study, two statistical models are proposed and compared with human allocation preferences in a subsequent validation study.

For the scenarios and budget sizes used, the results indicate that both the scenario ($F(2, 360) = 18.423$, $p < 0.05$) and budget ($F(4, 360) = 61,474$, $p < 0.05$) are statistically significant for smell. For vision, the budget was also significant ($F(4, 360) = 42.116$, $p < 0.05$) as was the scenario ($F(2, 360) = 11.032$, $p < 0.05$). For audio, the factor scenario was not found to be significant ($F(2, 360) = 0.0241$, $p > 0.05$) while the factor budget was statistically significant ($F(4, 360) = 153.85$, $p < 0.05$). Furthermore, for the scenarios where the smell source was found evident in the scene by the participants, the smell control was more times ON (see figure 10.4).

As far as the validation study is concerned, the model that is based on both the budget and scenario type ($\mathbf{M_2^{E2}}$) performs better than the model that is based solely on the budget ($\mathbf{M_1^{E2}}$) across all the test budget sizes and test scenarios.

11.4 Limitations

The limitations of the work presented in this thesis are the following:

- The main methodological limitation of the experimental studies ***E1*** and ***E3*** is the static nature of the visual stimuli. It is interesting to verify whether these results hold in animations or interactive VEs. The experimental framework can be extended to include videos instead of images; however, the use of sliders for watching different quality video clips could be more tedious and time consuming for the participants. The choice of rendered images provided further control over other parameters that could affect participants' decisions with videos such as frame rate, compression algorithm, video-audio synchronisation etc. Also with images, complications with long rendering hours can be avoided.

- The use of one smell only (citral) is another methodological limitation that could be improved in future studies. The use of many different odours could also increase the level of realism. This presupposes also the availability of an olfactory display that can support multiple odour channels along with a control channel.

- The use of normalised costs can also be considered a limitation. The alternative of working with physical computation times depends on the algorithm

selection for rendering visual and auditory stimuli. For physically based rendering using ray tracing the computation of an image can take several orders of magnitude compared with a RIR.

- Another limitation is the absence of task dependency in the experimental framework. The assigment of a task to the participants can affect allocation preferences depending on the nature of the task and the quality of the modality that is required for performing it.

11.5 Future Work

There are many ways this work can be extended in the future. This section provides possible insights and directions for further investigation of the problem of resource allocation presented in this thesis.

11.5.1 The impact of the Scenario context

The effect of the scenario content on participants' choices was clear in the experimental studies presented in this thesis. People clearly showed a preference to some scenarios more than others due to the nature of the auditory, visual or olfactory cues present in these VEs. Future work should investigate the effect of the scenario type more strictly. Broader classification of the scenarios similar to the ones found in this thesis would significantly assist the development of more accurate estimation models for resource allocation in multi-sensory VEs. One potential way to investigate the importance of the scenario type further is through examining whether visual and/or auditory salient features in the scene are able to affect participants' allocation strategy. Scenarios that contain conspicuous features are more likely to get assigned more computational resources either in the visual or auditory domain.

In real life situations, the sensation, perception and cognition of various sensory stimuli is not a process that happens in the same way at all times but changes dynamically. The distribution of resources based on human preference criteria is also a process that varies over time. Future studies should further investigate temporally varying resource allocation schemas based on the users' level of habituation and/or the scenario content.

11.5.2 The effect of the task in distributing resources

The top down visual attention is a perceptual mechanism that has been successfully exploited for reducing rendering times in selective rendering frameworks. Task maps

have been used to encode the objects of a virtual scene that are associated with a task that has been assigned to a viewer [CCW03]. It is an interesting question to investigate how human preferences on resource allocation change when a task is assigned to a user. The task can be, for example, to locate an object in the virtual scene or detect a token of an audio stream (e.g. a name). The modality appropriateness hypothesis supports that the sense that is more appropriate for a task will dominate the overall perception of the task. The quality of the displayed stimulus will definitely determine whether people are able to accomplish the task successfully but the extra quality that should be given to the most "appropriate" sense is something that should be investigated experimentally.

11.5.3 HDR imagery

HDR images and video have been proven to better approach the way the human eyes receive visual cues. The results of this thesis should be validated using HDR content instead of conventional LDR imagery. In terms of hardware there is not, to the best of our knowledge, a monitor that can display HDR content in ultra high definition (4K). The methodology presented in this thesis can be applied to repeat the experiments with HDR content.

11.5.4 Multidisciplinary approaches

Another perspective to the same problem could be to use methods from neuroscience for investigating the importance of each sense at any given time by studying brain activity at different regions of the head. It is not clear, of course, whether the experimental apparatus that is used in these areas can bias people's sensitivity to multi-sensory stimuli and how the collected signals can be correlated with human selection preferences.

11.5.5 Real-time rendering

Another possible direction for future work is to validate the results of this study with a real time renderer and not with pre-computed audio and visual stimuli as implemented in this thesis. This includes many challenges as computing and streaming physically-based rendered images in real time is still unachievable with common desktop computers used today.

11.6 Final Remarks

The simulation and delivery of high fidelity sensory cues have extended the limits of realism even further in the last few years. The ongoing development of even better hardware and software can contribute to the aim of creating virtual experiences equivalent to real life situations but this cannot be achieved without taking the human factor into account. VEs are constructed with the purpose to serve human needs and it would be an antithesis not to use human perceptual criteria in their design, computation and delivery.

The purpose of this thesis was to contribute to the growing area of perceptually based rendering by investigating optimal ways of distributing computational resources to the senses that are considered more important depending on the scenario and the available budget for quality improvements. This work can be the basis for rendering pipelines that take into account quality preferences for load balancing and computation of the VE stimulations.

Chapter 12

Acronyms

- **2IFC:** Two Interval Forced Choice
- **3D:** Three-Dimensions
- **AIC:** Akaike Information Criterion
- **ANOVA:** Analysis of Variance
- **BRDF:** Bi-directional reflectance distribution function
- **CFD:** Computational Fluid Dynamics
- **CIE:** Commission Internationale d'Eclairage
- **CSF:** Contrast Sensitivity Function
- **CV:** Control Volume
- **DSP:** Digital Signal Processing
- **FHD:** Full High Definition
- **GC-MS:** Gas Chromatography - Mass Spectometry
- **GUI:** Graphical User Interface
- **HAS:** Human Auditory System
- **HD:** High Definition
- **HDR:** High Dynamic Range
- **HOS:** Human Olfactory System

- **HSS:** Human Sensory System
- **HVS:** Human Visual System
- **JND:** Just Noticeable Difference
- **LDR:** Low Dynamic Range
- **LED:** Light emitting diode
- **LTI:** Linear and Time invariant
- **MANOVA:** Multivariate Analysis of Variance
- **PDF:** Probability Density Function
- **PID:** Photo Ionization Detector
- **QHD:** Quarter High Definition
- **RIR:** Room Impulse Response
- **UHD:** Ultra High Definition
- **VDP:** Visual Differences Predictor
- **VE:** Virtual Environment
- **VR:** Virtual Reality

Appendix A

Solid angle

A solid angle, denoted as Ω, is the 3D analogue of an angle as it is defined in the 2D Euclidean Geometry. A solid angle subtended from an object at a point $\mathbf{x} \in \mathbb{R}^3$ which is located at the center of a hemisphere is defined as the area S that is projected on the unit hemisphere. The projected area S can be given by tracing the silhouette of the object with a line that originates from the point $\mathbf{x}$ and scans the circumference of the distant object (see Figure A.1).

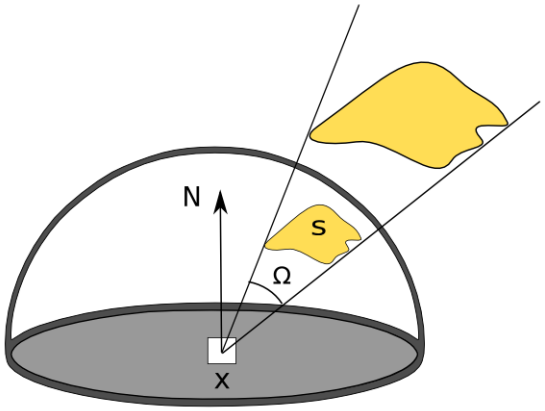


Figure A.1: A solid angle subtended at a point $\mathbf{x} \in \mathbb{R}^3$ by a distant object.

Solid angles are measured in *Steradians* (sr) and generally can be computed as:

$$\Omega = \frac{S}{r^2} \tag{A.1}$$

where S is the area projected on the hemisphere and r is the hemisphere's radius. Solid angles can be intuitively understood as a scale of how small or large a distant object seems to be from the standpoint of an observer located at point $\mathbf{x}$. As a

result, objects that are located further away form the point x will subtend smaller solid angles. The total solid angle of a unit hemisphere is 2π while the solid angle of a unit sphere is 4π steradians. A differential surface element $\mathrm{d}A$ that is at distance d from the point $\mathbf{x}$ subtends solid angle equal with:

$$\mathrm{d}\Omega = \frac{\mathrm{d}A \cos\theta}{d^2} \tag{A.2}$$

where θ is the angle between the normal vector of the differential area and the conceivable line that connects the differential area with point $\mathbf{x}$.

Appendix B

LTI filters and Convolution

The process of Convolution is of main importance in the context of Digital Signal Processing (DSP) applications as it used to express the output $y[n]$ of a Linear and Time Invariant filter (LTI) when its impulse response is known. The filter denoted as $\mathcal{H}$ has the properties of linearity and time invariance when:

$$\begin{aligned} &\mathcal{H}\{\alpha x_1[n] + \beta x_2[n]\} = \alpha\mathcal{H}\{x_1[n]\} + \beta\mathcal{H}\{x_2[n]\} \quad \textit{Linearity property} \\ &y[n] = \mathcal{H}\{x[n]\} \Leftrightarrow \mathcal{H}\{x[n-n_0]\} = y[n-n_0] \qquad \textit{Time Invariance} \end{aligned} \tag{B.1}$$

If the input to an LTI filter is a delta function $\delta[n]$ the output $\mathcal{H}\{\delta[n]\}$ is a signal that fully characterises the LTI system and is known as the impulse response of the filter. If the impulse response, denoted as $h_[n]$, is known for a filter then it is possible to find the output of any signal $x[n]$ that is given as input to the filter $\mathcal{H}$ using an operation known as convolution. Specifically, the filtered output $y[n]$ will be given as:

$$y[n] = x[n] \otimes h[n] = \sum_{k=-\infty}^{+\infty} x[k]h[n-k] \tag{B.2}$$

Convolution is an operation that is commutative and associative, In other words:

$$\begin{aligned} &(x \otimes h)[n] = (h \otimes x) = \sum_{k=-\infty}^{\infty} h[k]x[n-k] \quad \textit{Commutativity} \\ &\big((x \otimes h) \otimes w\big)[n] = \big(x \otimes (h \otimes w)\big)[n] \qquad \textit{Associativity} \end{aligned} \tag{B.3}$$

In the context of this thesis, convolution is used for sound spatialisation where the room geometry is the LTI filter $\mathcal{H}$ and the impulse response is used to introduce delays and attenuations to an anechoic sound stream. The impulse response includes all these reverberation effects.

Appendix C

Elements of vector calculus

Basic differential operators

The *gradient* operator ∇ is defined as follows:

$$\nabla = \frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k = \Big(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \Big)^T \tag{C.1}$$

The application of the gradient operator to a scalar function $f : (x, y, z) \mapsto \mathbb{R}$ gives the vector:

$$\nabla f = \Big(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \Big) \tag{C.2}$$

where x, y, z are the function's spatial coordinates. The application of the gradient operator to a vector valued function, for example, the function $\mathbf{U} : (x, y, z) \mapsto \big(u(x, y, z), v(x, y, z,), w(x, y, z) \big)$ gives a matrix as follows:

$$\nabla \mathbf{U} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{bmatrix} \tag{C.3}$$

The dot product of the gradient operator with a vector valued function is termed the *divergence* of the function and is denoted as $\nabla \bullet \mathbf{U}$ or $\mathrm{div}(\mathbf{U})$. It is defined as:

$$\mathrm{div}(\mathbf{U}) = \nabla \bullet \mathbf{U} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \tag{C.4}$$

The divergence of the gradient is known as the *Laplacian* operator and is denoted as Δ. For a scalar valued function f it is given as:

$$\mathrm{div}(\nabla f) = \nabla \bullet \nabla f = \nabla^2 f = \Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \tag{C.5}$$

The Laplacian of a vector valued function $\mathbf{U}$ is defined in the same way as a scalar valued function for each of its components:

$$\mathrm{div}(\nabla \mathbf{U}) = \nabla \bullet \nabla \mathbf{U} = \nabla^2 \mathbf{U} = \Delta \mathbf{U} = \begin{bmatrix} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \\ \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \\ \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \end{bmatrix} \tag{C.6}$$

Taylor expansion

The Taylor expansion of a function $g : \mathbb{R}^3 \mapsto \mathbb{R}$ around a point $\mathbf{a} = (a_1, a_2, a_3) \in \mathbb{R}^3$ can be written as follows:

$$g(\mathbf{x}) = g(\mathbf{a}) + (\mathbf{x} - \mathbf{a}) \cdot \left(\nabla g\right)_{\mathbf{a}} + \frac{1}{2!}(\mathbf{x} - \mathbf{a})^2 : \left(\nabla^2 g\right)_{\mathbf{a}} + \frac{1}{3!}(\mathbf{x} - \mathbf{a})^3 :: \left(\nabla^3 g\right)_{\mathbf{a}} + \cdots \tag{C.7}$$

where the operator : denotes array multimplication in this particular case. A more analytical expression for the first few terms is given as follows:

$$\begin{aligned} (\mathbf{x} - \mathbf{a}) \cdot \left(\nabla g\right)_{\mathbf{a}} &= \sum_{i=1}^{3} \frac{\partial g(\mathbf{a})}{\partial x_i}(x_i - a_i) \\ (\mathbf{x} - \mathbf{a})^2 : \left(\nabla^2 g\right)_{\mathbf{a}} &= \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{\partial^2}{\partial x_i \partial x_j} g(\mathbf{a})(x_i - a_i)(x_j - a_j) \\ &\vdots \end{aligned} \tag{C.8}$$

In the simpler case where a scalar valued function g is given, the Taylor expansion around a point $a \in \mathbb{R}$ takes the form:

$$g(x) = g(a) + \frac{g'(a)}{1!}(x - a) + \frac{g''(a)}{2!}(x - a)^2 + \frac{g'''(a)}{3!}(x - a)^3 + \cdots \tag{C.9}$$

Appendix D

Ethics information sheets

This appendix outlines the ethics considerations taken for the experimental studies ***E1***, ***E2***, ***E3*** of this thesis. One application has been made for all the three studies at Biomedical & Scientific Research Ethics Comittee (BSREC). Section D.1 presents the Participant Information Letter (PIL) that informs the participants about the overall procedure and objectives of the study. It covers sections about expenses, data and participant confidentiality, possible problems, etc. Section D.2 presents the invitation email and section D.3 includes the ethical approval letter for all the studies of this thesis.

D.1 Participant Information leaflet

THE UNIVERSITY OF WARWICK

PARTICIPANT INFORMATION LEAFLET

Study Title: Investigation of Cross-modal interactions and resources allocation at multi-sensory environments.

Investigator(s): Efstratios Doukakis (Doctoral Researcher) - WMG

Introduction

You are invited to take part in a research study. Before you decide, you need to understand why the research is being done and what it would involve for you. Please take the time to read the following information carefully. Talk to others about the study if you wish.

(Part 1 tells you the purpose of the study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study)

Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

PART 1

What is the study about?

Multi-sensory Virtual Environments (VE) can provide the experience of being in a simulated environment where multiple sensory stimuli (visual, auditory, olfactory, etc.) are presented at the same time. VEs have many potential uses in applications ranging from education and training to design and prototyping and they have also been successfully used in treating cognitive disabilities in healthcare. The construction and use of high fidelity VEs presupposes many computational resources (computers), a fact that noticeably limits their usage.

This study takes advantage of a basic psychological principle: People perceive the environment using all their available sensory organs at the same time. Also it is true that human sensory systems can contribute certain resources for identifying and perceiving environment information. When trying to construct virtual experiences where the quality level exceeds human's ability to perceive them, the extra computational resources are completely wasted.

The objective of this study is to investigate how people tend to allocate a given amount of computational resources so as to get quality improvements to two or more sensory stimuli.

Do I have to take part?

It is entirely up to you to decide. We will describe the study and go through this information sheet, which we will give you to keep. If you choose to participate, we will ask you to sign a consent form to confirm that you have agreed to take part. You will be free to withdraw at any time, without giving a reason and this will not affect you or your circumstances in any way.

1

What will happen to me if I take part?

At this experiment you will be presented with visual, auditory or/and olfactory cues at the same time. The experimental set up includes two monitors, a set of headphones and depending on the number of senses, a device that release smell bursts. At the first monitor images of different quality will be presented while at the second monitor a graphical user interface will include sliders bars (thumbs) that will be used to increase\decrease the quality of the delivered visual\auditory stimulus. The task is to adjust the sliders in such a way that the two delivered stimuli form the best virtual experience.

A tutor will indicate you all the necessary details; where to sit in front of a display, wear sound headphones or place the head near a smell outlet tube. Before the start of the experiment, a short tutorial will be given so as the user can familiarise with the graphical user interface.

What are the possible disadvantages, side effects, risks, and/or discomforts of taking part in this study?

As it was mentioned previously this study includes stimulus presentation using visual, auditory and if the experiment includes three senses, olfactory stimuli at the same time. This section describes possible risks or side effects for participating to this experiment.

As far as the presented images are concerned, the distance from the display will be sufficient not to make you feel dizzy or tired.

Furthermore, it should be made clear that the delivered sound is not going to be loud and quality improvements in this sense are not related at all with the sound intensity.

The released smells will be in low intensity and the smell will not be offensive.

It could also be said that none of the presented images or sounds will have unpleasant or offensive content. We have selected the scenarios in such a way that the vast majority of people will be familiar with. If for any reason any of the applied stimuli make you feel sick or tired, you will be accompanied to the room next to the laboratory for precautionary reasons.

What are the possible benefits of taking part in this study?

The data of this experiment will be used in our research to reduce the computational cost of constructing virtual experiences with little or no loss of fidelity. Your participation to this research study is a contribution to knowledge and science.

It is of great interest to us and other potential researchers to learn how to optimally partition a given amount of computational resources between the involved senses (vision, audition). The objective, as pointed previously, is to create a virtual environment perceptually indistinguishable from the one that all the senses were computed at the same quality level.

Expenses and payments

Participation in this research is all about voluntarily contributing to the current knowledge and there is no compensation for your expenses.

What will happen when the study ends?

After the end of the experiment the collected data will be used to a statistical analysis to determine whether multiple senses can actually improve identification and at what computational cost. We are interested in studying how these identification costs are distributed for each of the involved senses and make decisions for their level of importance. We will inform you about any achievement related to your participation.

Will my taking part be kept confidential?

Yes. We will follow strict ethical and legal practice and all information about you will be handled in confidence. Further details are included in Part 2.

What if there is a problem?

Any complaint about the way you have been dealt with during the study or any possible harm that you might suffer will be addressed. Detailed information is given in Part 2.

This concludes Part 1.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

PART 2

Who is organising and funding the study?

This study is fully financed by the Engineering and Physical Sciences Research Council (EPSRC). This study is part of a wider research project, organised by Jaguar Land Rovers in collaboration with the Warwick Manufacturing Group. This research project includes the creation and testing of a new driving simulator.

What will happen if I don't want to carry on being part of the study?

Participation in this study is entirely voluntary. Refusal to participate will not affect you in any way. If you decide to take part in the study, you will need to sign a consent form, which states that you have given your consent to participate.

If you agree to participate, you may nevertheless withdraw from the study at any time without affecting you in any way.

You have the right to withdraw from the study completely and decline any further contact by the staff after you withdraw.

If you agree to participate and you are student, your withdrawal will not affect any of your academic and professional records.

What if there is a problem?

This study is covered by the University of Warwick's insurance and indemnity cover. If you have an issue, please contact the Chief investigator of the study:

Efstratios Doukakis,

Doctoral Researcher
Warwick Manufacturing Group (WMG),
University of Warwick,
Coventry,
CV4 7AL
Email: E.Doukakis@warwick.ac.uk
Phone: 024765 73650 (ext. 73650)

Who should I contact if I wish to make a complaint?

Any complaint about the way you have been dealt with during the study or any possible harm you might have suffered will be addressed. Please address your complaint to the person below, who is a Senior University of Warwick official entirely independent of this study:

Director of Delivery Assurance
Registrar's Office
University House
University of Warwick
Coventry
CV4 8UW
Complaints@Warwick.ac.uk
024 7657 4774

Will my taking part be kept confidential?

The collected data will be captured anonymously, so there will be no confidential data (e.g. name, age, sex, etc.) from any participants in our database. Also, at the end of the experiment all the digital data will be written in a notepad and locked to investigator's office.

What will happen to the results of the study?

The research will be published at conference or academic journals related to this topic. Results will also be appeared in the thesis once the doctoral project is complete. The research outcomes will be used as a comprehensive guideline for researchers/developers who are working in the Cross-modal rendering domain.

Who has reviewed the study?

This study has been reviewed and given favourable opinion by the University of Warwick's Biomedical and Scientific Research Ethics Committee (BSREC): REGO-2014-625, date: 31st March 2014

What if I want more information about the study?

If you have any questions about any aspect of the study or your participation in it not answered by this participant information leaflet, please contact:

Doctoral Researcher:
Efstratios Doukakis ,
Warwick manufacturing Group,
University of Warwick,
Coventry, CV4 7AL,
Email: E.Doukakis@warwick.ac.uk
Phone: 02476573650 (ext. 73650)

Supervisors:
Professor Alan Chalmers,

Professor of Visualisation and Royal Society Industrial Fellow,
Warwick manufacturing Group,
University of Warwick,
Coventry, CV4 7AL,
Email: Alan.Chalmers@warwick.ac.uk
Phone: 02476524307

Dr Kurt Debattista,
Associate Professor in Visualisation Group,
Warwick Manufacturing Group (WMG),
University of Warwick,
Coventry, CV4 7AL,
Email: K.Debattista@warwick.ac.uk

Thank you for taking the time to read this participant information leaflet.

D.2 Invitation Email

Invitation email:

Title: Investigation of Cross-modal interactions and resources allocation at multi-sensory environments.

You are being invited to take part in a research study. Before you decide whether to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully.

The construction of immersive virtual environments includes the simulation of many different senses (vision, audition, olfaction, haptics, etc.) simultaneously. For the time being, this is possible only when many computational resources are available. This study will investigate how humans tend to allocate a budget of resources to the senses of vision, audio smell.

You will be presented with a graphical interface that includes two sliders. Each slider adjusts the quality of the delivered stimuli and your task will be to allocate your resources in order to get a satisfactory virtual experience.

Accepting to take part in this study is entirely on a voluntary basis and you can choose to withdraw from the study at any point before or after the study has commenced without giving any reasons and with no consequences. If you accept to take part it will take up about 20 minutes of your time.

The data from the study will be recorded on a private database. All of the data collected from the study will comply with the Data Protection Act. It will be completely anonymised and treated as confidential.

The research will be published at peer-reviewed international conferences and/or through academic journals related to this topic. Results will also appear in a thesis once the doctoral project is complete.

If you require any further information, please do not hesitate to contact me at:

Researcher: Efstratios Doukakis,
Doctoral Researcher
International Manufacturing Centre,
University of Warwick,
Coventry,
CV4 7AL
Email: E.Doukakis@warwick.ac.uk
Phone: 024765 73650 (ext. 73650)

If you agree to the study you must opt-in by signing the consent form.

Thank you for taking time to read this information sheet.

D.3 Ethics approval letter

1st April 2015

Warwick Medical School

PRIVATE
Mr Efstrarios Doukakis
PhD Student
WMG
University of Warwick
Coventry
CV4 7AL

Dear Mr Doukakis,

Study Title and BSREC Reference: Investigation of Cross-modal interactions and resources allocation at multi-sensory environments. REGO-2014-625

Thank you for submitting your revised amendment to the above-named project to the University of Warwick's Biomedical and Scientific Research Ethics Sub-Committee for approval.

I am pleased to confirm that approval is granted and your study may commence.

Please keep a copy of the signed version of this letter with your study documentation.

Yours sincerely

P.P. Graham Hewitt

Professor Scott Weich
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific Research Ethics Sub-Committee
A010 Medical School Building
Warwick Medical School,
Coventry, CV4 7AL.
Tel: 02476-528207
Email: BSREC@Warwick.ac.uk

Medical School Building
The University of Warwick
Coventry CV4 7AL United Kingdom
Tel: +44 (0)24 7657 4880
Fax: +44 (0)24 7652 8375

THE UNIVERSITY OF WARWICK

Bibliography

[AA00] C. Alain and S.R. Arnott. Selectively attending to auditory objects. *Frontiers in Bioscience*, 1(5):202–212, 2000.

[AAR72] A.D. Allport, B. Antonis, and P. Reynolds. On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, 24(2):225–235, 1972.

[AB79] J.B. Allen and D.A. Berkley. Image method for efficiently simulating smallroom acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, 1979.

[AB04] D. Alais and D. Burr. The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3):257–262, February 2004.

[ACB+07] M. Akeroyd, J. Chambers, D. Bullock, A. Palmer, Q. Summerfield, P. Nelson, and S. Gatehouse. The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics. *Acoustic Society of America*, 121(2):1056–1069, 2007.

[ADP01] P. Atkins and J. De Paula. *Physical Chemistry, 7th Ed.* W. H. Freeman, 2001.

[ADTA01] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano. The Cipic hrtf database. *Proceedings 2001 IEEE workshop on Applications of Signal Processing to Audio and Electroacoustics*, pages 99–102, 2001.

[AG98] E. Acheson and M. Gamble. The applicability of Weber's Law to Smell. *The American Journal of Psychology*, 10(1):82–142, 1898.

[AMB06] D. Alais, C. Morrone, and D. Burr. Separate attentional resources for vision and audition. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1592):1339–1345, 2006.

[App68] A. Appel. Some techniques forshading machine renderings of solids. In *In AFIPS 68': Proceedings of the April 30–May 2, 1968, spring joint computer conference*, pages 37–45. ACM, 1968.

[AR96] Jeffrey A.S. and Robert R.B. Characterization of the primary photochemical events in bacteriorhodopsin and rhodopsin. In A.G. Lee, editor, *Rhodopsin and G-protein Linked Receptors*, volume 2 of *Biomembranes: A Multi-Volume Treatise*, pages 33 – 139. JAI, 1996.

[AS71] R C. Atkinson and R.M. Shiffrin. The Control Processes of Short-term Memory. Technical report, Stanford University, Stanford, California, 1971.

[ASA12] ASA. Scientists tuning in to how you tune out noise. *ScienceDaily, Acoustical Society of America*, 2012.

[AY05] B.W. Ache and J.M. Young. Olfaction: Diverse species, conserved principles. *Neuron*, 48(3):417–430, 2005.

[Aya15] U. Ayachit. The ParaView Guide: A Parallel Visualization Application. Technical report, Kitware, 2015.

[BC10] B. Brkic and A. Chalmers. Virtual smell: Authentic smell diffusion in virtual environments. In *Proceedings of the 7th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '10, pages 45–52, New York, NY, USA, 2010. ACM.

[BCB+09] B.R. Brkic, A. Chalmers, K. Boulanger, S. Pattanaik, and J. Covington. Cross-modal affects of smell on the real-time rendering of grass. In *Proceedings of the 25th Spring Conference on Computer Graphics*, SCCG '09, pages 161–166, New York, NY, USA, 2009. ACM.

[BD95] W. Barfield and E. Danas. Comments on the use of olfactory displays for virtual environments. *Presence*, 5(1):109–121, 1995.

[BDM+05] M. Bertram, E. Deines, J. Mohring, J. Jegorovs, and H. Hagen. Phonon tracing for auralization and visualization of sound. In *VIS 05. IEEE Visualization, 2005.*, pages 151–158, Oct 2005.

[BDW80] F.O. Bartell, E.L. Dereniak, and W.L. Wolfe. The theory and measurement of bidirectional reflectance distribution function (brdf) and bidirectional transmittance distribution function (btdf). In *Proceedings of SPIE, Vol 257 Radiation Scattering in Optical Systems*, pages 154–160, 1980.

[BH98] A.W. Bronkhorst and T. Houtgast. Auditory distance perception in rooms. *Nature*, 397:517–520, 1998.

[Bia16] M. Biarnes. List of Ionization Potentials for Common VOCs. Technical report, E-Instruments, 2016.

[BK15] M. Bannon and F. Kaputa. The newton–laplace equation and speed of sound. *Thermal Jackets, online repository*, 2015.

[BKT86] K.R. Boff, L. Kaufman, and J.P. Thomas. *Handbook of Perception and Human Performance, Vol. 1: Sensory Processes and Perception*. Wiley-Interscience, 1986.

[Bla96] J. Blauert. *Spatial Hearing: Psychophysics of Human Sound Localization*. MIT Press, 1996.

[BM13] B. Belmudez and S. Möller. Audiovisual quality integration for interactive communications. *EURASIP Journal on Audio, Speech, and Music Processing*, 2013(1):1–23, 2013.

[BMVK14] C. Bushdid, M.O. Magnasco, L.B. Vosshall, and A. Keller. Humans can discriminate more than 1 trillion olfactory stimuli. *Science*, 343(6177):1370–1372, 2014.

[Bor84] J. Borish. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6):1827–1836, 1984.

[BP80] V.A. Benignus and J.D. Prah. A computer-controlled vapor-dilution olfactometer. *Behavior Research Methods & Instrumentation*, 12(5):535–540, 1980.

[BP00] K. Brandenburg and H. Popp. An introduction to MPEG Layer-3. Technical report, Fraunhofer IInstitution, 2000.

[BR81] P. Bertelson and M. Radeau. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, 29(6):578–584, 1981.

[BR97] C.P. Brown and Duda R.O. An efficient HRTF model for 3D sound. In *Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on*, pages 1–4, Oct 1997.

[Bre94] A.S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, 1994.

[Bro27] R. Brown. A brief account of microscopical observations made in the months of June. July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine*, 4:161–173, 1827.

[Bro58] D.E. Broadbent. *Perception and Communication, 1st Ed.* Pergammon Press, London, 1958.

[BS04] E.R. Berg and G.D. Stork. *The Physics of Sound, 3rd ed.* Paerson, 2004.

[Cat04] K. Cater. *Detail to Attention: Exploting Limits of the Human Visual System for Selective Rendering*. PhD thesis, University of Bristol, 2004.

[CCL02] K. Cater, A. Chalmers, and P. Ledda. Selective quality rendering by exploiting human inattentional blindness: Looking but not seeing. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '02, pages 17–24, New York, NY, USA, 2002. ACM.

[CCTB11] M.M. Chun, J.D. Colomb, and N.B. Turke-Browne. A taxonomy of external and internal attention. *Annual Review of Psychology*, 62:73–101, 2011.

[CCW03] K. Cater, A Chalmers, and G. Ward. In Eurographics Symposium on Rendering 2003. In *Detail to Attention: Exploiting visual tasks for selective rendering*, pages 270–280, 2003.

[CDB09] A. Chalmers, K. Debattista, and B. Brkic. Towards high-fidelity multi-sensory virtual environments. *Visual Computer*, Vol.25(No.12):1101–1108, December 2009.

[CEN03] CEN. European Standard on Determination of Odour Concentration by Dynamic Olfactometry. Technical report, EN 13725, 2003.

[CF10] A. Chalmers and A. Ferko. Levels of realism: From virtual reality to real virtuality. In *Proceedings of the 24th Spring Conference on Computer Graphics*, SCCG '08, pages 19–25, New York, NY, USA, 2010. ACM.

[Che53] C.E. Cherry. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5):975–979, 1953.

[Che06] Y. Chen. Olfactory display: Development and application in virtual reality therapy. In *Artificial Reality and Telexistence–Workshops, 2006. ICAT '06. 16th International Conference on*, pages 580–584, Nov 2006.

[CHM09] A. Chalmers, D. Howard, and C. Moir. Real virtuality: A step change from virtual reality. In *Proceedings of the 25th Spring Conference on Computer Graphics*, SCCG '09, pages 9–16, New York, NY, USA, 2009. ACM.

[CIE24] *Commission Internationale de l'eclairage proceedings*, 1924.

[Cit16] Pubchem compound database. Technical report, National Center for Biotechnology Information, 2016.

[CMA+10] D. Chaudhury, L. Manella, A. Arellanos, O. Escanilla, T.A. Cleland, and C. Linster. Olfactory bulb habituation to odor stimuli. *Behavioural Neuroscience*, 124(4):490–499, 2010.

[Col74] F.B. Colavita. Human sensory dominance. *Perception and Phychophysics*, 16(2):409–412, 1974.

[Coo86] R.L. Cook. Stochastic sampling in computer graphics. *ACM Trans. Graph.*, 5(1):51–72, 1986.

[CPC84] R.L. Cook, T. Porter, and L. Carpenter. Distributed ray tracing. *SIGGRAPH Comput. Graph.*, 18(3):137–145, 1984.

[CSS04] G. Calvert, C. Spence, and B.E. Stein. *The Handbook of Multisensory Processes*. MIT Press, 2004.

[Cus97] E.L. Cussler. *Diffusion: Mass Transfer in Fluid Systems, 2nd Edition*. Cambridge University Press, 1997.

[CWGJ75] C.S. Choe, R.B. Welch, R.M. Gilford, and J.F. Juola. The "ventriloquist effect": Visual dominance or response bias? *Perception & Psychophysics*, 18(1):55–60, 1975.

[Dal92] S.J. Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. volume 1666, pages 2–15, 1992.

[Dar08] C.J. Darwin. Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):1011–1021, 2008.

[DB74] J.A. Desor and G.K. Beauchamp. The human capacity to transmit olfactory information. *Perception & Psychophysics*, 16(3):551–556, 1974.

[DBB06] P. Dutrè, K. Bala, and P. Bekaert. *Advanced Global Illumination, 2nd ed.* AK Peters, 2006.

[DC05] K. Debattista and A. Chalmers. Component-based adaptive sampling. In *XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'05)*, pages 375–382, 2005.

[DD95] R. Desimone and J. Duncan. Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, 18:193–222, 1995.

[DD06] B. David and A. David. Combining visual and auditory information. *Progress in Brain Research*, 155:243–258, 2006.

[Deb06] K. Debattista. *Selective Rendering for High Fidelity Graphics*. PhD thesis, University of Bristol, 2006.

[DFMM99] G. Dickins, M. Flax, A. McKeag, and D. McGrath. Proceeding of the AES 16th international conference, spatial sound reproduction, Rovaniemi, Finland. In *Optimal 3D–speaker panning.*, pages 421–426, 1999.

[DK98] J. Degel and E.P. Köster. Implicit memory for odors: a possible method for observation. *Perceptual and motor skills*, 86(3):943–952, 1998.

[DS94] J. Driver and C. Spence. *Attention and Performance XV*. MIT Press, 1994.

[DS98] J. Driver and C. Spence. Crossmodal attentionn. *Current Opinion in Neurobiology*, 8(2):245–253, 1998.

[Duf99] A. Dufour. Importance of attentional mechanisms in audiovisual links. *Experimental Brain Research*, 126(2):215–222, 1999.

[Dur94] B.R. Durand. *3-D Sound for Virtual Reality and Multimedia Applications*. Academic Press Inc, 1994.

[Dus92] B.D. Dusenbury. *Sensory Ecology: How organisms Acquire and Respond to Information*. W.H.Freeman & Co Ltd, 1992.

[Ear89] J. Eargle. Stereophonic Techniques – an anthology of reprinted articles on stereophonic techniques. *Audio Engineering Society*, 1989.

[Ern82] T.T. Ernst. *Fragrance chemistry: The science of sense of smell.* Academic Press, 1982.

[Est76] K.W. Estes. *Handbook of Learning and Cognitive Processes: Vol. 4, Attention and Memory.* Psychology Press, 1976. page 305.

[Etc80] M. Etcheto. A constant flow rate olfactometer with seven calibrated concentrations. *Chemical Senses*, 5(1):1–9, 1980.

[Fer01] J. Ferwerda. Elements of early vision for computer graphics. *IEEE Computer Graphics and Applications*, 21(5):22–33, Sep 2001.

[FMT11] N. Fouladinejad, Abd M.K., and J.T. Taib. Modeling virtual driving environment for a driving simulator. In *Control System, Computing and Engineering (ICCSCE), 2011 IEEE International Conference on*, pages 27–32, Nov 2011.

[FN96] W. Fujisaki and S. Nishida. Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Experimental Brain Research*, 166(3–4):455–464, 1996.

[For99] S. Fortune. Topological beam tracing. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry*, SCG '99, pages 59–68, New York, NY, USA, 1999. ACM.

[For13] K. Forinash. *Sound: An Interactive eBook.* Paerson, 2013.

[Fou27] J. Fourier. Mémoir sur les températures du globe terrestre et des espaces planétaires. *Mémoires de l'Académie Royale des Sciences*, 7:569–604, 1827.

[FP02] J.H. Ferziger and Milovan P. *Computational Methods for Fluid Dynamics.* Springer, 3rd Ed, 2002.

[FSQM03] C.A. Filho, C.M. Silva, M.B. Quadri, and E.A. Macedo. Tracer diffusion coefficients of citral and d-limonene in supercritical carbon dioxide. *Fluid Phase Equilibria*, 204(1):65–73, 2003.

[FTC+04] T. Funkhouser, N. Tsingos, I. Carlbrom, G. Elko, M. Sondhi, J.E. West, G. Pingali, P. Min, and A. Ngan. A beam tracing method for interactive atchitectural acoustics. *The Journal of Acoustic Society of America*, 115(2):739–756, 2004.

[GA97] R. Gilkey and T.R. Anderson. *Binaural and Spatial Hearing in Real and Virtual Environments*. Psychology Press, 1997.

[Gav12] S. Gavriel. *Handbook of Human Factors and Ergonomics*. Wiley, 4 edition, 2012.

[GBW+09] D. Grelaud, N. Bonneel, M. Wimmer, M. Asselot, and G. Drettakis. Efficient and practical audio-visual rendering for games using crossmodal perception. In *In I3D 2009: Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pages 177–182. ACM, 2009.

[GD03] J.A. Gottfried and R.J. Dolan. The nose smells what the eye sees: Crossmodal visual facilitation of human olfactory perception. *Neuron*, 39(2):375–386, 2003.

[Gel09] S. Gelfand. *Hearing: An Introduction to Psychological and Physiological Acoustics*. CRC Press, 5th Edition, 2009.

[Ger99] J. Gerstmann. Quake 3 Arena Review. Technical report, CBS Interactive Inc., 1999.

[Gle15] E. Glenn. *The Physics Hypertextbook*. Online Physics textbook, 2015.

[GLT05] E. Gallo, G. Lemaitre, and N. Tsingos. Prioritising signals for selective real-time audio processing. In *In Proceedings of Int. Conf. on Auditory Display*, ICAD, 2005.

[GM94] B. Gardner and K. Martin. HRTF measurements of a Kemar dummy-head microphone. *MIT Media Lab*, May 1994.

[GRK96] A.N. Gilbert, M. Robyn, and S.E. Kemp. Cross-Modal Correspondence between Vision and Olfaction. *The Americal Journal of Psychology*, 109(3):335–351, 1996.

[GTS+99] D.P. Greenberg, K.E. Torrance, P. Shirley, J. Arvo, E. Lafortune, J. A. Ferwerda, B. Walter, B. Trumbore, S. Pattanaik, and S. Foo. A framework for realistic image synthesis. *Commun. ACM*, 42(8):44–53, August 1999.

[Gue90] B.D. Guenther. *Modern Optics*. John Wiley & Sons, 1990.

[HAC08] V. Hulusic, M. Aranha, and A. Chalmers. The influence of cross-modal interaction on perceived rendering quality thresholds. In V. Skala, editor, *In WSCG 2008 Full Papers Proceedings*, pages 41–48, 2008.

[HB96] C.M. Hendrix and W. Barfield. Presence within virtual environments as a function of visual display parameters. *Presence*, 5(3):274–289, 1996.

[HBRDC11] C. Harvey, T. Bashford-Rogers, K. Debattista, and A. Chalmers. Visual saliency for smell impulses and application to selective rendering. In *Theory and Practice of Computer Graphics 2011*, pages 73–80. Eurographics Association, 2011.

[HCD+09] V. Hulusić, G. Czanner, K. Debattista, E. Sikudova, P. Dubla, and A. Chalmers. Investigation of the beat rate effect on frame rate for animated content. In *Proceedings of the 25th Spring Conference on Computer Graphics*, SCCG '09, pages 151–159, New York, NY, USA, 2009. ACM.

[HCG+11] J. Hidalgo, G. Chopard, J. Galmiche, L. Jacquot, and G. Brand. Just noticeable difference in olfaction: a discriminative tool between healthy elderly and patients with cognitive disorders associated with dementia. *Rhinology*, 49(5):513–518, Dec 2011.

[HCL03] D. Harel, L. Carmel, and D. Lancet. Towards an odor communication system. *Computational Biology and Chemistry*, 27(2):121–133, 2003.

[HDAC10] V. Hulusić, K. Debattista, V. Aggarwal, and A. Chalmers. Exploiting audio-visual cross-modal interaction to reduce computational requirements in interactive environments. In Kurt Debattista, editor, *2nd International Conference on Games and Virtual Worlds for Serious Applications VS-GAMES 2010 : proceedings : 25-26 March 2010, Braga Portugal*. IEEE Computer Society, Los Alamitos, Calif., 2010.

[HDAC11] V. Hulusić, K. Debattista, V. Aggarwal, and A. Chalmers. Maintaining frame rate perception in interactive environments by exploiting audio-visual cross-modal interaction. *The Visual Computer*, Volume 27(Number 1):57–66, January 2011.

[HDBRC16] C. Harvey, K. Debattista, T. Bashford-Rogers, and A. Chalmers. Multi-modal perception for selective rendering. *Computer Graphics Forum*, 2016.

[HE96] R.S. Herz and T. Engen. Odor Memory: Review and analysis. *Psychonomic Bulletin & Review*, 3(3):300–313, 1996.

[HEL68] R. Harper, C.B. Edgar, and D.G. Land. *Odour Description and Odour Classification: A multidisciplinary examination*. Harcourt Brace/Churchill Livingstone, 1968.

[Hel15] E. Heller. *Why you hear what you hear*. Online Psychoacoustics textbook, 2015.

[Hen16] H. Henning. *Der Geruch*. Leipzig, Germany: Barth, 1916.

[HHD+12] V. Hulusic, C. Harvey, K. Debattista, N. Tsingos, S. Walker, D. Howard, and A. Chalmers. Acoustic rendering and auditory–visual cross-modal perception and interaction. *Computer Graphics Forum*, 31(1):102–131, February 2012.

[HMYS01] J. Haber, K. Myszkowski, H. Yamauchi, and H.P. Seidel. Perceptually guided corrective splatting. In *In Proceeding of Eurographics*, pages 142–152, 2001.

[Hon13] Honeywell. The PID Handbook: Theory and Applications of Direct-Reading Photoionization Detectors (PIDs). Technical report, RAE Systems, 2013.

[HP08] J.M. Henderson and G.L. Pierce. Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, 15(3):566–573, 2008.

[HPI14] J.C. Hay, H.L. Pick, and K. Ikeda. Visual capture produced by prism spectacles. *Psychonomic Science*, 2(1):215–216, 2014.

[HRW13] D. Halliday, R. Resnick, and J. Walker. *Fundamentals of Physics Extended, 10th ed.* Wiley, 2013.

[HS13] K.S. Helly and K.M. Stanney. *Handbook of Virtual Environments 2nd Ed.* CRC Press, 2013.

[HSN12] National Health System (NHS). Overview of Anosmia. Technical report, NHS Choices, 2012.

[HT66] I.P. Howard and W.B. Templeton. *Human spatial orientation [by] I.P. Howard and W.B. Templeton*. Wiley, London, New York,, 1966.

[Hun04] R.G.W. Hunt. *The reproduction of Colour*. John Wiley & Sons, 2004.

[HWBR+10] C. Harvey, S. Walker, T. Bashford-Rogers, K. Debattista, and A. Chalmers. The Effect of Discretised and Fully Converged Spatialised Sound on Directional Attention and Distraction. In John Collomosse and Ian Grimstead, editors, *Theory and Practice of Computer Graphics*. The Eurographics Association, 2010.

[HZZ01] E. Hecht, A. Zajac, and A.W. Zajic. *Optics, 4th ed.* Addison-Wesley Publishing Company, 2001.

[IK99] L. Itti and C. Koch. A comparison of feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging*, 10:161–169, 1999.

[IK00] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Visual Research*, 40(10–12):1489–1506, 2000.

[IKN98] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1254–1259, 1998.

[ITU12] General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. ITU-R, August 2012.

[Jam80] W. James. *The principles of Psychology, Part 1.* Henry Holt and Company, 1980.

[Jas96] H. Jasak. *Error Analysis and Estimation for the Finite Volume Method with Applications to Fluid Flows.* PhD thesis, Imerial College, Department of Mechanical Engineering, 1996.

[JCT95] J. Jiang, P. Coffey, and B. Toohey. Improvement of odor intensity measurement using dynamic olfactometry. *Journal of the air & Waste management association*, 56(5):675–683, 1995.

[JJS93] N. Jayant, J. Johnston, and R. Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10):1385–1422, Oct 1993.

[JL72] W.P. Jones and B.E. Launder. The Prediction of Laminarization with a two-equation model of turbulence. *International Journal of Heat and Mass Transfer*, 15:301–314, 1972.

[JLW95] L. Jot, V. Larcher, and O. Warufsel. Digital signal processing issues in the context of binaural and transaural stereophony. In *Proccedings of the 98th Audio Engineering Society Convection*, 1995.

[KA88] D. Kirk and J. Arvo. The Ray Tracing Kernel. In *In Proceedings of Ausgraph*, pages 75–82, 1988.

[Kaj86] J.T. Kajiya. The rendering equation. *SIGGRAPH Comput. Graph.*, 20(4):143–150, 1986.

[KBM08] B. Kollmeier, T. Brand, and B. Meyer. *Springer Handbook of Speech Processing*, chapter Perception of Speech and Sound, pages 61–82. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[Kel11] A. Keller. Attention and olfactory consciousness. *Frontiers in Psychology*, 2(380):1–13, 2011.

[KN07] O. Kalinli and S.S. Narayanan. A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech. In *Proceedings of InterSpeech*, pages 1941–1944, Antwerp, Belgium, 2007.

[Kna95] S.C. Knasko. Pleasant odors and conguency: Effects on approach behaviour. *Chemical Senses*, 20:479–487, 1995.

[Kob85] G. Kobal. Pain-related electrical potentials of the human nasal mucosa elicited by chemical stimulation. *Pain*, 22(2):151–163, 1985.

[KP09] F. Kingdom and N. Prins. *Psychophysics: A Practical Introduction*. Academic Press, 2009.

[KPK04] M.A. Köster, J. Prescott, and E.P. Köster. Incidental learning and memory for three basic tastes in food. *Chemical Senses*, 29(5):441–453, 2004.

[KPLL05] C. Kayser, C. Petkov, M. Lippert, and N. Logothetis. Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology*, 15(8):1943–1947, 2005.

[KSS68] A. Krokstad, S. Strom, and S. Sørsdal. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125, 1968.

[KU85] C. Koch and S. Ullman. Shifts in selective attention: towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985.

[Kut09] H. Kuttruff. *Room Acoustics, 5th Edition*. CRC Press, 2009.

[L.93] Hilmar L. Systematic errors of the ray-tracing algorithm. *Applied Acoustics*, 38(2):207–221, 1993.

[LB92] H. Lehnert and J. Blauert. Principles of binaural room simulation. *Applied Acoustics*, 36(3):259–291, 1992.

[LBHM99] D. Lloyd, S. Bolanowski, L. Howard, and F. McGlone. Mechanisms of attention in touch. *somatosensory and & Motor research*, 16(1):3–10, 1999.

[LDC06] P. Longhurst, K. Debattista, and A. Chalmers. A GPU based saliency map for high-fidelity selective rendering. In *Proceedings of the 4th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '06, pages 21–29, New York, NY, USA, 2006. ACM.

[Lew93] T. Lewers. A combined beam tracing and radiatn exchange computer model of room acoustics. *Applied Acoustics*, 38(2):161–178, 1993.

[LEZP99] T.S. Lorig, D.G. Elmes, D.H. Zald, and J.V. Pardo. A computer-controlled olfactometer for fmri and electrophysiological studies of olfaction. *Behavior Research Methods, Instruments, & Computers*, 31(2):370–375, 1999.

[LFR04] J. Lachter, K. Forster, and E. Ruthruff. Forty-five years after Broadbent (1958): Still no identification without perception. *Psychology Review*, 111(4):880–913, 2004.

[LGA+10] J.N. Lundström, A.R. Gordon, E.C. Alden, S. Boesveldt, and J. Albrecht. Methods for building an inexpensive computer-controlled olfactometer for temporally-precise experiments. 78(2):179–189, 2010.

[LL98] A. Livermore and D.G. Laing. The influence of chemical complexity on the perception of multicomponent odor mixtures. *Perception & Psychophysics*, 60(4):650–661, 1998.

[LL06] S.B. Lowen and S.E. Lukas. A low-cost, MR-compatible olfactometer. *Behavior Research Methods*, 38(2):307–313, 2006.

[LMC92] M.T. Landahl and E. Mollo-Cristiansen. *Turbulence and Random Processes in Fluid Mechanics*. Cambridge University Press, 1992.

[LMLCBT06] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):802–817, May 2006.

[LNL06] R.H.Y. Leung, Braasch N.M, and K.L. Leung. A low cost, non-individualized surround system based upon head-related transfer functions. an ergonomics study and prototype development. *applied Ergonomics*, 37:695–707, 2006.

[LVK02] P. Larsson, D. Vastfjall, and M. Kleiner. Better presence and performance in virtual environments by improved binaural sound rendering. *Proceedings of the AES 22nd Int Conf. on virtual, synthetic and entertainment audio, Espoo, Finland*, 12(2):156–167, 2002.

[Mas05] G. Mastoropoulou. *The Effect of Audio on the Visual Perception of High-Fidelity Animated 3D Computer Graphics, PhD in Computer Science.* PhD thesis, University of Bristol, 2005.

[Mat10] H. Matsukura. Fluid dynamic considerations for realistic odor presentation using olfactory display. *Presence*, 19(6):513–526, Dec 2010.

[MBD01] G. Morrot, F. Brochet, and Denis D. The color of odors. *Brain and Language*, 79(2):309–320, 2001.

[MBM02] J. Millot, G. Brand, and N. Morand. Effects of ambient odors on reaction time in humans. *Neuroscience Letters*, 322(2):79–82, 2002.

[MBT+07] T. Moeck, N. Bonneel, N. Tsingos, G. Drettakis, I. Viaud-Delmon, and D. Aloza. Progressive perceptual audio rendering of complex scenes. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. ACM SIGGRAPH, April 2007.

[MC80] C. Murphy and S.W. Cain. Taste and olfaction: Independence vs interaction. *Physiology & Behavior*, 24(3):601–605, 1980.

[MC97] A.S. Mandayam and B. Cagatay. Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers & Graphics*, 21(4):393–404, 1997.

[MC04] G. Mastoropoulou and A. Chalmers. The effect of music on the perception of display rate and duration of animated sequences: an experimental study.

In *Theory and Practice of Computer Graphics, 2004. Proceedings*, pages 128–134, June 2004.

[MC15] F.N. Martin and J.G. Clark. *Introduction to Audiology, 12th Ed.* Pearson, 2015.

[MDCT05a] G. Mastoropoulou, K. Debattista, A. Chalmers, and T. Troscianko. Auditory bias of visual attention for perceptually-guided selective rendering of animations. In *Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, GRAPHITE '05, pages 363–369, New York, NY, USA, 2005. ACM.

[MDCT05b] G. Mastoropoulou, K. Debattista, A. Chalmers, and T. Troscianko. The influence of sound effects on the perceived smoothness of rendered animations. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*, APGV '05, pages 9–15, New York, NY, USA, 2005. ACM.

[MDMS05] R. Mantiuk, J.S. Daly, K. Myszkowski, and H. Seidel. Predicting visible differences in high dynamic range images - model and its calibration. In Bernice E. Rogowitz, Thrasyvoulos N. Pappas, and Scott J. Daly, editors, *Human Vision and Electronic Imaging X, IS&T/SPIE's 17th Annual Symposium on Electronic Imaging (2005)*, volume 5666, pages 204–214, 2005.

[MKRH11] R. Mantiuk, K.J. Kim, A.G. Rempel, and W. Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, 30(4):40:1–40:14, July 2011.

[MM72] T.R. Marrero and E.A. Mason. Gaseous diffusion coefficients. *Journal of Physical and Chemical Reference Data*, 1(1):3–118, 1972.

[MM76] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264(5588):746–748, 1976.

[Møl89] H. Møller. Reproduction of artificial-head recordings through loudspeakers. *Audio Engineering Society*, 37:30–33, February 1989.

[Møl92] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36:171–218, 1992.

[Moo03] B.C.J. Moore. *An introduction to Psychology of Hearing.* Academic Press, 5th ed., 2003.

[Mor04] F. Morganti. Virtual interaction in cognitive neuropsychology. *Studies in health technology and informatics*, 99(1):55–70, 2004.

[MR98] A. Mack and I. Rock. *Inattentional Blindness*. MIT Press, 1998.

[MT94] D.A. McCormick and B. Thierry. Sensory gating mechanisms of the thalamus. *Current Opinion in Neurobiology*, 4(4):550–556, 1994.

[MW77] D.W. Massaro and D.S. Warner. Dividing attention between auditory and visual perception. *Perception & Psychophysics*, 21(6):569–574, 1977.

[MW01] G.F. Meyer and S.M. Wuerger. Cross-modal integration of auditory and visual motion signals. *Neuroreport*, 8(12):2557–2560, 2001.

[Nag80] T. Nagano. Temporal sensitivity of the human visual system to sinusoidal gratings. *Journal of the Optical Society of America*, 70(6):711–716, 1980.

[Nef98] P. Nef. The molecular and cellular bases of olfaction. *News in Physiological Sciences*, 13(1):1–5, 1998.

[Nic09] A.J. Nicell. Assessment and regulation of odour impacts. *Atmospheric Environment*, 43(1):196–206, 2009. Atmospheric Environment - Fifty Years of Endeavour.

[Nie92] S.H. Nielsen. Auditory Distance Perception in Different Rooms. Mar 1992.

[NIS01] National institute of standards and technology (nist). Technical report, Gaithersburg, Maryland, U.S.A, 1901.

[NMDSLC15] M. Narwaria, R.K. Mantiuk, M.P. Da Silva, and P. Le Callet. HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501, 2015.

[NW72] J. Nelder and R.W. Wedderburn. Generalized Linear Models. *Journal of the Royal Statistical Society*, 135(3):370–384, 1972.

[PAF01] D. Purves, A.J. Augustine, and D. Fitzpatrick. *The organization of the Olfactory System*. Sinauer Associates, 2001.

[Pas99] H. Pashler. *The psychology of attention*. The MIT Press, 1999.

[PfSI12] Programme for Simulation and Innovation. Jaguar Land Rovers and EPSRC. Technical report, Collaboration Project, 2012.

[PH10] M. Pharr and G. Humphreys. *Physically Based Rendering, Second Edition: From Theory To Implementation.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2010.

[PIW11] M.H. Pinson, W. Ingram, and A. Webster. Audiovisual quality components. *IEEE Signal Processing Magazime*, 28(6):60–67, 2011.

[PKT+08] D. Pang, A. Kimura, T. Takeuchi, J. Yamato, and K. Kashino. A stochastic model of selective visual attention with a dynamic bayesian network. In *2008 IEEE International Conference on Multimedia and Expo*, pages 1073–1076, June 2008.

[PNK76] M.I. Posner, M.J. Nissen, and R.M. Klein. Visual dominance: an information-processing account of its origins and significance. *Psychological review*, 83(2):157, 1976.

[Pos80] M.I. Posner. Orienting of Attention. *The quarterly journal of Experimental Psychology*, 32(1):3–25, 1980.

[PS90] D.R. Perrott and K. Saberi. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, 87(4):1728–1731, 1990.

[PSNG06] C.T. Pearce, S.S. Schiffman, H.T. Nagle, and W.J. Gardner. *Handbook of Machine Olfaction: Electronic Nose Technology.* Wiley-VCH, 2006.

[Pug88] E. Pugh. *Vision: Physics and Retina Physiology, Steven's Handbook of Experimental Psychology, 2nd Ed.* John Wiley & Sons, 1988.

[PUJG02] R. Patterson, S. Uppenkamp, I. Johnsrude, and T. Griffiths. The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4):767–776, 2002.

[Ray07] R.S. Rayleigh. Xii. On our perception of sound direction. *Philosophical Magazine Series 6*, 13(74):214–232, 1907.

[RBvdZ+00] A. Rizzo, J.G. Buckwalter, C. van der Zaag, U. Neumann, M. Thiebaux, C. Chua, A. van Rooyen, L. Humphrey, and P. Larson. Virtual environment applications in clinical neuropsychology. In *Virtual Reality, 2000. Proceedings. IEEE*, pages 63–70, 2000.

[RCHR07] B. Ramic, A. Chalmers, J. Hasic, and S. Rizvic. Selective rendering in a multi-modal environment: Scent and graphics. In *Proceedings of the 23rd*

Spring Conference on Computer Graphics, SCCG '07, pages 147–151, New York, NY, USA, 2007. ACM.

[Rey86] *Phil. Transactions, Royal society of London*, 1886.

[RFWB07] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala. Visual equivalence: Towards a new standard for image fidelity. *ACM Trans. Graph.*, 26(3), July 2007.

[Rim10] E. Rimmel. *The Book of Perfumes*. Kessinger Publishing, 2010.

[Riv03] G. Riva. Applications of virtual environments n medicine. *Methods of Information in medicine*, 42(3):524–534, 2003.

[RLC+07] N. Raghuvanshi, C. Lauterbach, A. Chandak, D. Manocha, and Ming C. Lin. Real-time sound synthesis and propagation for games. *Commun. ACM*, 50(7):66–73, July 2007.

[RLGM92] I. Rock, C.M. LInnett, P. Grant, and A. Mack. Perception without attention: results of a new method. *Cognitive Psychology*, 24(4):502–534, 1992.

[RM15] E. Robert and R. Miller. *User Guide to VST plug-in Ambiophonics DSP*. On-Line Book, 2015.

[ROC97] R.A. Resnink, K.J. O'Regan, and J.J. Clark. To see or not see: The need for attention to perceive changes in scenes. *In Investigative Opthamology & Visual Science*, 8(5):368–373, 1997.

[Rog25] P.M. Roget. Explanation of an optical deception in the appearance of the spokes of a wheel seen through vertical apertures. *Philosophical Transactions of the Royal Society of London*, 115:131–140, 1825.

[RSD+02] C. Rouby, B. Schaal, D. Dubois, R. Gervais, and A. Holley. *Olfaction, Taste and Cognition*. Cambridge University Press, 2002.

[SAD07] P. Satarzadeh, R.V. Algazi, and R.O. Duda. Physical and filter pinna models based on anthropometry, 2007.

[SAM09] P. Shirley, M. Ashikhmin, and S. Marschner. *Fundamentals of Computer Graphics, 3rd ed.* AK Peters, 2009.

[SB78] Patankar S.V. and Baliga B.R. A new finite-difference scheme for parabolic differential equations. *Numerical Heat Transfer*, 1(1):27–37, 1978.

[SC99] D.J. Simmons and C.F. Chabris. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, 28(9):1059–1074, 1999.

[SD97] C. Spence and J. Driver. On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, 59(3):389–403, 1997.

[SD04] C. Spence and J. Driver. *Crossmodal Space and Crossmodal Attention*. Oxford University Press, 2004.

[SDJ+10] A. Seigneuric, K. Durand, T. Jiang, J. Baudouin, and B. Schaal. The nose tells it to the eyes: Crossmodal associations between olfaction and vision. *Perception*, 39(11):1541–1554, 2010.

[SDLC05] V. Sundstedt, K. Debattista, P. Longhurst, and A. Chalmers. Visual attention for efficient high-fidelity graphics. In *In SCCG 05: Proceedings of the 21st spring conference on Computer Graphics*, pages 169–175. ACM Press, 2005.

[SE98] R. Streicher and F. Everest. *The new stereo soundbook, 2nd ed.* Audio Engineering Associate, 1998.

[SHHT86] L. Savioja, J. Huopaniemi, T. Huotilainen, and T. Takala. Real-time virtual audio reality, 1986.

[SHS+04] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Trans. Graph.*, 23(3):760–768, August 2004.

[SHTV99] L. Savioja, J. Huopaniemi, Lokki T., and R. Väänänen. Creating Interactive Virtual Acoustic Environments. *AES E-Library*, 47(9):675–705, 1999.

[SIS+05] N. Sakai, S. Imada, S. Saito, T. Kobayakawa, and Y. Deguchi. The effect of visual images on perception of odors. *Chemical Senses*, 30(suppl 1):244–245, 2005.

[SJ01] D.L. Strayer and W.A. Johnston. Driven to distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone. *Psychological Science*, 12(6):462–466, 2001.

[SKS00] L. Shams, Y. Kamitani, and S. Shimojo. Illusions: What you see is what you hear. *Nature*, 408:788, 2000.

[SLKS07] S. Siltanen, T. Lokki, S. Kiminki, and L. Savioja. The room acoustic rendering equation. *The Journal of the Acoustic Society of America*, 122(3):1624–1632, 2007.

[SM93] B.E. Stein and M.A. Meredith. *The merging of the senses.* The MIT Press, 1993.

[SMKK01] C. Spence, F.P. McGlone, B. Kettenmann, and G. Kobal. Attention to olfaction. A psychophysical investigation. *Experimental Brain Research*, 138(4):432–437, 2001.

[Smy97] J. Smythies. The functional neuroanatomy of awareness: With a focus on the role of various anatomical systems in the control of intermodal attention. *Consciousness and Cognition*, 6(4):455–481, 1997.

[SNH+10] R.H.S. So, B. Ngan, A. Horner, J. Braasch, and K.L. Leung. Toward orthogonal non-individualized head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study. *Ergonomics*, 53(6):767–781, 2010.

[SNS99] C. Scheier, R. Nijhawan, and S. Shimojo. Sound alters visual temporal resolution. *Investigative Ophthalmology & Visual Science*, 40(Suppl. 4):4169, 1999.

[SP04] S.S. Schiffman and T.C. Pearce. Introduction to Olfaction: Perception, Anatomy, Physiology and Molecular Biology. *Wiley-VCH Verlag GmbH and Co, KGaA*, pages 1–31, 2004.

[SR14] P.C. Sande and S. Ray. Mesh size effect on CFD simulation of gas-fluidized geldart a particles. *Powder Technology*, 264:43–53, 2014.

[SRD00] C. Spence, J. Ranson, and J. Driver. Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention. *Perception & Psychophysics*, 62(2):410–424, 2000.

[SRMN10] H. Seo, E. Roidl, F. Müller, and S. Negoias. Odors enhance visual attention to congruent objects. *Appetite*, 54(3):544–549, 2010.

[Ste88] E.I. Stearns. Commission internationale de l'clairage, colorimetry, 2nd ed. *Color Research & Application*, 13(1):64–65, 1988.

[Ste01] J.R. Stevenson. Associative learning and odor quality perception: How sniffing an odor mixture can alter the smell of its parts. *Learning and Motivation*, 32(2):154–177, 2001.

[Ste11] S.W. Steven. *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, 2011.

[Sto98] R.L. Storms. *Auditory-Visual Cross-Modal Perception Phenomena*. PhD thesis, Naal Postgraduate School, 1998.

[SW55] R.A. Schneider and S. Wolf. Olfactory perception thresholds for citral utilizing a new type olfactorium. *Journal of Applied Physiology*, 8(3):337–342, 1955.

[TC73] W.R. Thurlow and J.E. Charles. Certain Determinants of the "Ventriloquism Effect". *Perceptual and Motor Skills*, 36(3):1171–1184, 1973.

[Tel92] S. Teller. *Visibility Computations in*. PhD thesis, Computer Science Department, University of California, Berkeley, 1992.

[TGD04] N. Tsingos, E. Gallo, and G. Drettakis. Perceptual audio rendering of complex virtual environments. *ACM Trans. Graph.*, 23(3):249–258, August 2004.

[THWA+95] Ss Tsan-Hsing, W.L. William, S. Aamir, Y. Zhigang, and Z. Jiang. A new k-ϵ eddy viscosity model for high reynolds number turbulent flows. *Computers & Fluids*, 24(3):227 – 238, 1995.

[TN03a] D.J. Tellinghuisen and E.J. Nowak. The inability to ignore auditory distractors as a function of visual task perceptual load. *Perception & Psychophysics*, 65(5):817–828, 2003.

[TN03b] D.J. Tellinghuisen and E.J. Nowak. The inability to ignore auditory distractors as a function of visual task perceptual load. *Perception & Psychophysics*, 65(5):817–828, 2003.

[Tor09] A. Torralba. How many pixels make an image? *Visual Neuroscience*, 26(1):123–131, 2009.

[TS10] J.M. Taekman and K. Shelley. Virtual environments in healthcare: immersion, disruption, and flow. *International Anesthesiology clinics*, 48(3):101–121, 2010.

[Tsi05] N. Tsingos. Scalable Perceptual Mixing and Filtering of Audio Signals using an Augmented Spectral Representation. In *8th International Conference on Digital Audio Effects (DAFx 2005)*, page 6, Madrid, Spain, 2005.

[Vir16] ARCH Virtual. Virtual Reality Development. Technical report, Private Company, 2016.

[VM07] H. Versteeg and W. Malalasekera. *An Introduction to computational fluid dynamics: The finite volume method*. Prentice Hall, 2nd Ed., 2007.

[VS13] L.R. Varshney and J.Z. Sun. Why do we perceive logarithmically? *Significance*, 10(1):28–31, 2013.

[VWVO04] M. Van Wanrooij and A. Van Opstal. Contribution of head shadow and pinna cues to chronic monaural sound localization. *Journal of Neuroscience*, 24(17):4163–4171, 2004.

[WA01] E.H. White and Jenkins A.F. *Fundamentals of Optics*. McGraw-Hill Primis Custom Publishing, 2001.

[Wal02] G. Waldman. *Introduction to Light: The Physics of Light, Vision, and Color*. Dover Publications, 2002.

[Wat90] M. Watt. Light-water interaction using backward beam tracing. *SIGGRAPH Computer Graphics*, 24(4):377–385, September 1990.

[Whi80] T. Whitted. An improved illumination model for shaded display. *Commun. ACM*, 23(6):343–349, 1980.

[WOC00] P.M. Wise, M.J. Olsson, and W.S. Cain. Quantification of odor quality. *Chemical Senses*, 24(4):429–443, 2000.

[Wol99] J.M. Wolfe. Inattentional Amnesia. Fleeting Memories. *Cognition of Brief Visual Stimuli*, pages 71–94, 1999.

[WR93] A. Wolfgang and F. Rainer. EARS Auralization Software. *Audio Engineering Society Convention (AES)*, 11(2):894–904, 1993.

[Wri81] D.W. Wright. 50 years of the 1931 CIE standard observer. *DIE Farbe*, 29(4), 1981.

[WSY+12] T. Weiss, K. Snitz, A. Yablonka, R.M. Khan, D. Gafsou, E. Schneidman, and N. Sobel. Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white. 109(49):19959–19964, 2012.

[WTJF98] H.G. Weller, G. Tabor, H. Jasak, and C. Fureby. A tensorial approach to computational continuum mechanics using object-oriented techniques. *Comput. Phys.*, 12(6):620–631, November 1998.

[WW80] R.B. Welch and D.H. Warren. Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3):638–667, 1980.

[WW08] R.D. Wright and L.M. Ward. *Orienting of Attention, 1st Edition*. OUP USA, 2008.

[WWP88] R.M. Warren, J.M. Wrightson, and J. Puretz. Illusory continuity of tonal and infratonal periodic sound. *The Journal of the Acoustical Society of America*, 84(4):1338–1342, 1988.

[Yar67] A. Yarbus. Eye movements during perception of complex objects. *In L.A. Riggs, Ed., Eye Movements and Vision 7*, pages 171–196, 1967.

[YN06] K. Yoshikawa and T. Nakamoto. Movie with scents generated by olfactory display using solenoid valves. In *Virtual Reality Conference, 2006*, pages 291–292, March 2006.

[YPG01] H. Yee, S. Pattanaik, and D.P. Greenberg. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph.*, 20(1):39–65, 2001.

[YSYT94] T. Yasuda, T. Suzuki, S. Yokoi, and J. Toriwaki. Virtual environment construction for driving simulator. In *Robot and Human Communication, 1994. RO-MAN '94 Nagoya, Proceedings., 3rd IEEE International Workshop on*, pages 50–55, Jul 1994.

[Zwa95] H.H. Zwaardemaker. *Physiologie des Geruchs*. Leipzig, Verlag von Wilhelm Engelmann, 1895.